A system level approach to identify novel cell size regulators

Dul, Zoltan

Awarding institution:
King's College London

A system level approach to identify novel cell size regulators

University of London

ZOLTAN DUL

A THESIS PRESENTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

2019

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where stated otherwise by reference or acknowledgement, the work presented is entirely my own.

Zoltán Dul

January 2019

Acknowledgements

Thank you to my primary supervisor **Dr Attila Csikász-Nagy** for giving me the opportunity to work in his computational biology group in London and for guiding me so well throughout my PhD studies. His sound ideas and plans immeasurably helped my way to complete my PhD. I'm also exceptionally grateful to my second supervisor **Professor Shaun Thomas** for helping and introducing myself to wet lab experiments with T-Cells. His academic experience and inspirational ideas have been invaluable to me. I also would like to thank you, my Italian supervisor, **Dr Azeddine Si-Ammour** for guiding my path in Arabidopsis research and giving me research ideas. I'm very grateful to him introducing me to a new research area in science.

I would like to acknowledge the help of the Leukaemia Cell Cycle Lab members in the Rayne Institute, King's College London especially to **Dr Stephen Orr** who introduced me the various techniques of T-Cell experimentation and guided my path until I mastered the technique. Also, I'm thankful to **Dr Nermina Lamadema** and **Dr Sun Sook Chung** without whom this work would not have been possible to complete. Moreover, I'm indebted to **Dr Loretta Brown** and others at King's College Hospital for helping me in some areas of research.

I would like you to thank the current and former members and Csikasz-Nagy group: **Dr Federico Vaggi**, **Rosa Hernansaiz** and **Kirsten Jenkins** for their help in the Bioinformatic field. Moreover, I'm very grateful to **Dr Asquini Elisa**, **Dr Mirko Moser** and **Dr Valentina Cappelletti** and all other people from Fondazione Edmund Mach, San Michele all' Adige, Italy for guiding me in the endless field of Arabidopsis research.

Finally, I would like to thank the Research and Innovation Centre in Fondazione Edmund Mach, San Michele all' Adige, Italy for funding my PhD studies.

Last but not least to my family and friends, thank you for all the support and encouragement.

Abstract

It is known that evolutionarily conserved pathways regulate several essential biological processes, such as the cell cycle, DNA replication and protein synthesis. I hypothesise that there exists a conserved regulatory mechanism which controls cell size as well. To test this proposition, I carried out two types of experiment:

Bioinformatics. Development of a bioinformatics tool to predict cell size regulators, based on information of evolutionarily-conserved proteins which have been described as cell size regulators in genome-wide studies. I collected existing large-scale data from the literature from five evolutionarily distant organisms (*A. thaliana, D. melanogaster, H. sapiens, S. cerevisiae* and *S. pombe*), then looked for conserved orthologous proteins with conserved cell size regulatory functions. I used the eggNOG orthology database as the primary source for orthology information, in addition to one manually curated list by PomBase curators. I added biological pathway information from KEGG and functional data from Gene Ontology. This approach allowed me to identify a core conserved cell size regulatory network and based on data on three of these species I created a list of predicted novel cell size regulators for the remaining two organisms. I focussed on those orthologous groups of proteins, which have only one ortholog in all five organisms and further reduced the list by concentrating on those which lack pathway annotations. 20 conserved orthologous groups matched these criteria. Some of these were tested in wet-lab experiments.

Additionally, I have extended this tool with the capability to find the conserved core of any biological function, listed in the Gene Ontology SLIM database and extended the list of the investigated model species to seven (*A. thaliana, C. elegans, D. melanogaster, D. rerio, H. sapiens, S. cerevisiae* and *S. pombe*). This bioinformatics tool is also able to predict novel functional annotations to Gene Ontology, based on orthologous group involvement in eggNOG.

Wet-lab experiments. Analysis of some of the predicted cell size regulators by carrying out experiments on *Arabidopsis thaliana* and human T-Cells.

T-Cells. I isolated human peripheral blood T-Cells that are in a quiescent state, which increase in size and enter the cell cycle when stimulated *via* CD3/CD28. The quiescent T-Cells were transfected with siRNA to reduce the induction of each of the predicted proteins, which normally occurs as the cells respond to CD3/CD28. I analysed the size distribution by Flow Cytometry, measuring FSC-A values. My data show that reducing the induction of TCTP, a protein that plays a role in microtubule stabilisation, apoptosis and calcium binding, affects T-Cell size. I also show that reducing the expression of the MCM7 protein, a member of the mini-chromosome maintenance protein complex, reduced the size of T-Cells in G_1 and increased their size in the S-phase of the cell cycle.

Arabidopsis thaliana. I confirmed previously published data that knockout mutants of *pfd6*, a member of the prefoldin complex, produce small plants. I show that knock-out *pfd3* mutants have a significantly smaller cell size compared with the control. The knock-out *pfd3* mutants also have significantly different seed and silique sizes. Furthermore, I found that the knockout of *mos14*, encoding a nuclear import receptor, results in significantly smaller leaves.

In summary, I have created a flexible bioinformatics tool that can predict novel cell size regulators, some of which are verified by experiments on human cells and *Arabidopsis*.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

APCs	Antigen-presenting cells
ATP	Adenosine triphosphate
Cdk	Cyclin-dependent kinase
CFSE	Carboxyfluorescein succinimidyl ester
CKI	Cyclin-dependent kinase inhibitor
Col-0	Columbia type Arabidopsis plant
CSV	Comma Separated Value (File type)
DNA	Deoxyribonucleic acid
dpg	days post germination
ENSG	Ensembl Gene Identifier
FACS	Fluorescent activated cell sorters
FBS	Foetal Bovine Serum
FITC	Fluorescein Isothiocyanate
FNP	First neighbour protein (bioinformatics)
Foxp3	forkhead box P3
FSC	Forward Scatter (Flow Cytometry)
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GO	Gene Ontology Database
GOOT	Gene Ontology Orthology Tool (my tool)
ITAM	Immunoreceptor Tyrosine-based Activation Motif
KEOPS complex	Kinase, putative endopeptidase and other proteins of small size
KO	Knock-Out (gene)
LDS	Lithium Dodecyl Sulfate
MCM complex	Minichromosome maintenance protein complex
MGG	May-Grünwald Giemsa (staining)
MHC	Major histocompatibility complex
mTOR	Mammalian Target of Rapamycin

NCBI	The National Center for Biotechnology Information
NK cells	Natural Killer cells
NK-T cells	Natural-Killer T-Cells
PBMCs	Peripheral Blood Mononuclear Cells
PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
PE	Phycoerythrin
PI	Propidium Iodine
pMHC	Peptides associated with major histocompatibility complex class I
PPI	Protein-protein interaction
Rb	Retinoblastoma protein
RNA	Ribonucleic acid
SCP	Cell size control protein (bioinformatics)
SDS	Sodium Dodecyl Sulfate
SGD	Saccharomyces cerevisiae Genome Database
siRNA	Small interfering RNA
SSC	Side Scatter (Flow Cytometry)
SVG	Scalable Vector Graphics
TAIR	The Arabidopsis Information Resource
TCR	T-Cell Receptor
TCTP	Synonim for TPT1, as Translationally Controlled Tumour Protein
TOR	Target of Rapamycin
TPT1	Tumour Protein, Translationally-Controlled 1
UID	Uniprot ID
VBP1	Von Hippel-Lindau Binding Protein 1
VD	Venn Diagram
YAP	Yes associated protein

”Once you eliminate the impossible,
whatever remains,
no matter how improbable,
must be the truth”
Sir Arthur Conan Doyle

Chapter 1

General Introduction

1.1 Aims of PhD Dissertation

There are cellular processes that are conserved in many species. The work in my Thesis addresses a related fundamental question in biology:

Are there common regulators of cell size in evolutionarily distant organisms?

To answer this complex question, I will collect data from the academic literature manually, then use a self-created bioinformatic tool to predict evolutionarily conserved proteins that control cell size, some of which will be tested experimentally. During the data collection and bioinformatic analyses will check other bioinformatic tools such as the Gene Ontology and eggNOG databases as additional sources of information. I test my predicted candidates using two distinct organisms to determine whether they regulate cell size. I chose human T-Cells and *Arabidopsis thaliana*, which are evolutionarily distant organisms. Analyses of these species will also determine whether there are common cell size regulators in in mammals and plants.

The key aims of my Thesis are to:

- **Develop** a bioinformatics tool that predicts novel groups of genes involved in cell size control by using system-wide, peer-reviewed studies, ortholog databases and Gene Ontology annotations.
- **Identify** novel cell size regulator genes from some of those predicted by the bioinformatics tool by carrying out gene knockout experiments on the plant *Arabidopsis thaliana*.
- **Identify** novel cell size regulators in Human cells by using small interfering RNA to reduce the expression of some of the predicted genes in human peripheral blood T-Cells.
- **Expand** the bioinformatics tool with functionality to find the conserved core of a given functional annotation and give novel annotations for groups of proteins in a specifically selected Gene Ontology annotation based on their orthological relationship.

1.2 Cell size

1.2.1 Cell size difference and uniformity

In nature, cell sizes in different organisms vary enormously as do the sizes of cells within an organism. Despite the wide range of cell sizes, the size of individuals within each species is relatively uniform; this is true for overall size as well as for the dimensions and proportion of organs within the body. This kind of uniformity means strict control of cell, organ and organism size mediated through complex coordination of cellular growth and proliferation (Conlon & Raff, 1999; Cook & Tyers, 2007).

1.2.2 Cellularity

In unicellular species, cell size can vary by ten thousand-fold. The smallest known unicellular organism, *Mycoplasma genitalium* (Sippel et al., 2012), is around 0.1 μm in diameter, while the largest unicellular organisms candidate can be *Acetabularia* and *Gromia sphaerica* (Hammerling, 1963; Marshall et al., 2012; Matz et al., 2008), which are approximately 3 meters in diameter. Cell size in multicellular organisms also differs. The average size of an animal cell is between 10–20 μm in diameter (Guertin & Sabatini, 2006). In humans despite there being around 30 trillion cells in the body (Bianconi et al., 2013), there are around 200 cell types (Fu et al., 2017), all holds its own characteristic of size, which indicate precise regulatory mechanisms to control cell size (Guertin & Sabatini, 2006). Just to give some examples, how diverse a human cell can be: a human egg has a diameter of 0.1 mm, an axon of a human motoneuron can be over 1 meter long, while neutrophil blood cells are only 10 μm in diameter (Baserga, 2007; Guertin & Sabatini, 2006).

In a unicellular organism, the cell size determines the size of the organism; in a multicellular species, the combination of the number of cells and their sizes determines the organism size. Although in advanced multicellular organisms, cells form organs and the numbers of cells in the organ determines the size of that organ. However, the size

control of organs and tissues also affect the size of cells in the organ (Edgar, 2006). Master hormonal regulators such as insulin/insulin-like growth factor signalling hormones control the growth of cells, tissues and body of *Drosophila melanogaster* (Goberdhan & Wilson, 2003). Insulin pathway defects cause cellular growth defects (Böhni et al., 1999), while hyperactivation of the pathway causes overgrowth (Weinkove et al., 1999). Hyper nutrient usage and increased number of steroid hormone receptors can also affect the cell growth (Edgar, 2006).

Some perturbations of cell size, like a change in ploidy, can alter cell size, without modifying the organ size (Cook & Tyers, 2007; Jorgensen & Tyers, 2004). Generally, cell sizes scale with ploidy, this was shown for example in *S. cerevisiae,* where diploid cells are almost double the size of haploid cells (Weiss et al., 1975). However, this is not true for all the organisms. The most notable example of this is an experiment done by Fankhauser using salamander larvae (Fankhauser, 1945), where the ploidy was increased, but the size of the organs remained the same. This is the case for pentaploid animals that were grown under normal laboratory conditions also, where the number of cells was just one fifth that of the original salamanders (Fankhauser, 1945).

In other cases, the size of organs can be larger, without any changes in size at the cellular level. To give an example, in transgenic mouse studies, where *CDKN1B* (encodes $p27^{Kip1}$) deficient mice have enlarged organs due to increased cell proliferation, while cellular dimensions remain the same (Kiyokawa et al., 1996; Nakayama et al., 1996; Tumaneng et al., 2012). However, in most of the cases, cells "sense" the organ size and in the case of a mutation such as *CDK1* mutants in *Drosophila* (Su & O'Farrell, 1998; Weigmann et al., 1997), cells decrease in size to fit the dimensions of an organ. (Cook & Tyers, 2007).

1.2.2.1 Regulation of cell size

Evolutionarily conserved pathways regulate several essential biological processes in cells such as the cell cycle engine, which affects cell proliferation (Cross et al., 2011; van den Heuvel & Dyson, 2008) and core processes of ribosome biogenesis (Thomson et al., 2013). Since cell size is widely variable throughout the phylogenetic tree (Conlon & Raff, 1999), there could also be conserved mechanisms, which regulate cell size. These mechanisms, which control cell size and maintain cells within a particular size range have been investigated for a long time since the microscope was discovered by Robert Hooke (Hooke, 1665). However, the mechanisms that maintain a uniform cell size within a narrow range are poorly understood (Marshall et al., 2012).

Maintaining the size of actively dividing cells requires strict regulation. Normally, when stem cells receive mitogenic stimuli, they proliferate and produce daughter cells in confined size distribution (Yamamoto & Mak, 2017). However, there are particular examples where this is not the case. For example, mechanical injury to tissue can alter cell size and results in cells that are transiently bigger (Kim et al., 2006). On the other hand, asymmetric cell division can happen in species such as *C. elegans*, producing daughter cells that are different, as is the case when segregation of the anterior-posterior axis takes place (Goldstein & Hird, 1996). In normal asymmetric cell division the components of the mother cell, other than DNA are differentially segregated to each of the daughter cells, resulting in daughter cells that differ in size but with the same DNA content (Hawkins & Garriga, 1998).

The cell size in organisms generally reflects the balance between cell growth and division (Cook & Tyers, 2007; Ferrezuelo et al., 2012; Jorgensen & Tyers, 2004). Indeed these two processes maintain the size of cells in a particular range (Cook & Tyers, 2007; Jorgensen et al., 2004; Yamamoto et al., 2014). In this thesis I am looking for the conserved mechanisms that are regulating the size of eukaryotic cells.

1.2.3 Checkpoints in cell size control

Mutations or malfunctions can cause changes in growth rate that affect the cell size (Cook & Tyers, 2007; Rupes, 2002; Yamamoto et al., 2014). If a timer mechanism applies (see later in Section 1.3.2) and the growth rate accelerates, cells will be larger, while a slowing of growth may lead to smaller cells. Growth rate can directly determine the critical size[1] of cells (Aldea et al., 2017). There exists a so-called size checkpoint or size threshold (Cook & Tyers, 2007). This is, in most cases, at the point of cell cycle commitment, in yeast termed "*Start*"(Hartwell et al., 1970), while in mammalian cells the "*Restriction Point*" (Pardee, 1974; Zetterberg & Larsson, 1985). These points are that checkpoints in the cell cycle where cells commit to cell division. Cell-size checkpoints are regulated by the environmental conditions (the number of nutrients in the media) in yeasts (Nurse et al., 1976). Cells growing at a slower rate will have a smaller cell size when they are maintained in nutrient-poor media, while cells proliferating in nutrient-rich media usually grow at a faster rate and reach a larger cell size (Nurse et al., 1976). If they are switched from a nutrient-poor medium to a nutrient-rich medium (or *vice-versa*), the cells will adapt to the new conditions (Fantes & Nurse, 1977; Nurse et al., 1976). This behaviour has been shown to occur in mammalian Schwann cells as well (Conlon & Raff, 2003). It was shown that the growth rate of cells is also dependent on the current size of those cells (Tzur et al., 2009). Reportedly there exist size thresholds, which are dynamically set by growth factor signals in mammalian cells (Dolznig et al., 2004; Zetterberg & Killander, 1965). However, in some cases, it seems to be that cell size is a simple result of factors that influence growth rate (Conlon & Raff, 2003). It has been argued that the signalling pathway proteins that are downstream of growth factor receptors, such as Akt/PKB, MTOR, cyclin D-CDK4 and pRb, can also influence cell size (Cook & Tyers, 2007; Edgar, 2006). In Section 1.3.3.1 , I explain in detail how TOR related proteins influence cellular growth and cell size.

[1] A size that is required to be reached to proceed through one of the critical cell cycle transitions.

1.2.4 Coordination of cell growth and cell division

Cell size homeostasis in proliferating cells requires accurate coordination of cell division, such that on average each cell division is accompanied by a doubling of cell mass (Jorgensen & Tyers, 2004). In cells which are not undergoing mitotic cell division, such as neurons, the balance of general anabolic and catabolic reactions maintain cell size (Jorgensen & Tyers, 2004). To maintain cell size, cells must coordinate the processes of cell growth and the cell cycle, which depends on the model that the cells follow. If cells grow linearly then even with noise on the process, cell size will be maintained in a certain regime. Whereas if cells increase their size exponentially then larger cells can get larger in each cycle and small cells can get smaller in each cycle leading to a fitness loss (Conlon & Raff, 2003; Dolznig et al., 2004). Certain size control mechanisms are proposed to maintain the size of these cells in homeostasis by coupling cell cycle progression to reaching a critical size (see Section 1.3.2). These mechanisms are thought to be conserved and coupled processes for a long period, although studies in recent 20 years have proved they can be uncoupled under certain circumstances (Conlon et al., 2001; Lea, Orr, et al., 2003). For example in T-Cells our lab has identified a point, named "Commitment Point" that can uncouple the transition from G_0 to G_1 in T-Cells, thus regulate cell growth during a mitotic cycle (Lea, Orr, et al., 2003). Our lab has described that cells need to get external signals such as CD3 and CD28 for 3-5 hours to commit themselves to the cell cycle (Lea, Orr, et al., 2003). If the signal stimulus is removed earlier, cells are going to return the G_0 state, while if it removed later the cells going to progress in the cell cycle. This commitment point is dependent on the activation of CDK6/4 D Cyclin and can be inhibited by the cellular Cdk4/6-cyclin D inhibitor $p16^{INK4A}$ (see Section 3.1.2.4.3.1).

Cell division is usually dependent on cell growth, but these processes may also be regulated independently (see Figure 1.2-1). Blocks to cell cycle progression do not necessarily prevent growth and can result in abnormally sized daughter cells (Neufeld & Edgart, 1998; Neufeld et al., 1998). Although direct interactions between the regulators of cell cycle and cell growth suggest that the two processes may be controlled by partially overlapping networks in metazoans (Cook & Tyers, 2007). Below I discuss

examples of mechanisms that control cell cycle progression that are and are not evolutionarily conserved as well as mechanisms that regulate cell growth and to what extent these processes are coupled.

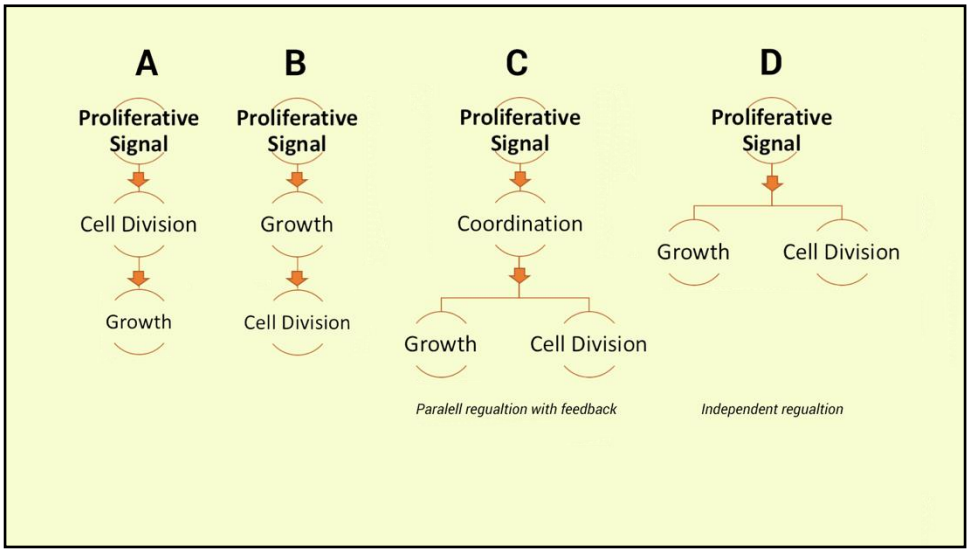


Figure 1.2-1: The four possible relationships between cell division and cell growth.

(A) Cell division drives growth.

(B) Growth drives cell division.

(C) Growth and cell division rates are controlled in parallel by a common upstream regulator.

(D) Independent regulation of cell division and growth

The figure is based on Neufeld et al. (1998).

1.3 Connection between cell size and the cell cycle

Different cell size regulation processes may exist in organisms in the phylogenetic tree separated by millions of years of evolution. Also, organisms such as yeasts and mammals share cell cycle regulatory machinery at the molecular level (Cook & Tyers, 2007; Rupes, 2002). However, there are differences in protein families regulating such processes, such as the cyclin-dependent kinases (CDK) (Malumbres & Barbacid, 2009; Risal et al., 2016), as described below.

1.3.1 Molecular regulators of cell size in yeasts

Two well-studied species, the budding yeast and the fission yeast, provide good genetic models in which to study cell-cycle and cell size control (Jorgensen et al., 2007; Rupes, 2002). These organisms are considered to grow exponentially, thus they require an active size control mechanism to maintain their size in a homeostatic regime. The central regulation of cell size consists of two different molecular processes. Each organism has a control point in G_1/S and G_2/M cell cycle phases, but the importance of these points are different (Rupes, 2002; Turner et al., 2012).

In fission yeast, the key regulator of cell size occurs in G_2/M (Jorgensen et al., 2007; Marshall et al., 2012; Rupes, 2002). Here the key role is played by Cdc2 (encoded by the Cell Division Cycle gene *CDC2)*, (Moreno et al., 1989; Simanis & Nurse, 1986), which regulates progression through the cell cycle. Cdc2 is inhibited by Wee1 (Russell & Nurse, 1987), which is inhibited by the complex of Cdr1/Cdr2 proteins (Martin & Berthelot-Grosjean, 2009; Moseley et al., 2009). A key regulator of cell size is Pom1, which is slowly released from the tips of the rod-shaped fission yeast cells. As the cell grows and elongates the relative concentration of Pom1 decreases in the cell, especially near the nucleus. Thus, it cannot inhibit Cdr1 anymore, which inactivates Wee1 leading to Cdc2 T14/Y15 dephosphorylation and Cdc2 activation of, inducing mitosis (see Figure 1.3-1).

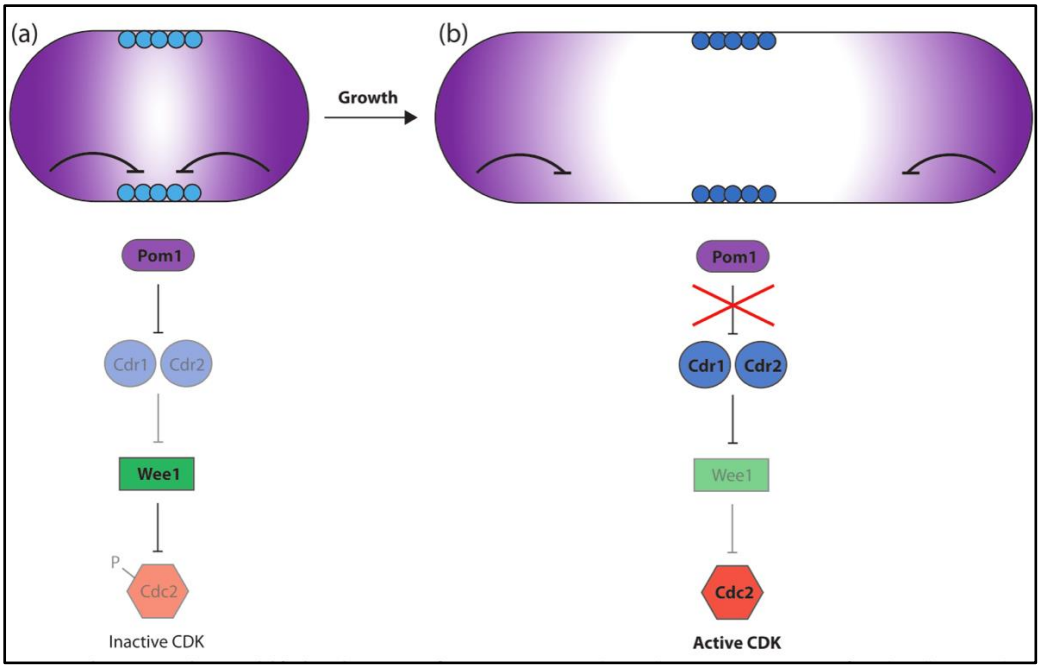


Figure 1.3-1: Cell size control in Fission yeast and Pom1 gradient model.

Details of Cdc2 regulation by Pom1 are described in the text.

The figure is reproduced from Marshall et al. (2012).

In budding yeast, the important checkpoint is at the G_1/S transition. A G_1 cyclin protein called Cln3 regulates "*Start*" (Marshall et al., 2012; Rupes, 2002; Turner et al., 2012) by inhibiting Whi5 and thereby allowing transcription of genes required for cell cycle progression (de Bruin et al., 2004). The rate of protein synthesis affects the level of Cln3 because of its high turnover rate, thereby coordinating cell growth with progression through the cell cycle (Marshall et al., 2012; Rupes, 2002).

It seems that there is a different primary regulation in distinct yeast species. However, there exist common regulators that affect the regulation of cell size. A good example is Sch9 protein from budding yeast; it is an S6 kinase, that is activated by the phosphorylation of TORC1 complex, while in response it inhibits Rim15, a protein that controls the entry into the G_0 (Pedruzzi et al., 2003). Sch9 has three different orthologs in fission yeast: Sck1, Sck2 and Psk1 (Nakashima et al., 2010), and recently it has been described one of them, Sck2 affects greatwall-endosulfine Prpk18 (Chica et al., 2016).

1.3.2 Different concepts of cell size control

In proliferating cells the cell size is determined by a balance between cell growth and cell division (Tzur et al., 2009).
There are three models in the literature that employ different mechanisms:

- **Sizer model**: In this case a cell actively monitors its size, then if it reaches a certain size (threshold size), it will divide (Facchetti et al., 2017). One widely known example for this concept is the Whi5-Start mechanism in budding yeast (Facchetti et al., 2017; Turner et al., 2012).

- **Timer model**: In the timer model, cells are in the growth phase for a certain amount of time, after which they enter the cell cycle. In fission yeast the amount of time spent in the G2 phase is according to this timer model (Sveiczer et al., 1996).

- **Adder model**: The newest concept incorporates the assumption that cells add a constant amount of cell mass to increase size that is unrelated to the time spent in the cell cycle or the size of the new-born cell. A good example of this concept is *E. coli* bacteria homeostasis (Taheri-Araghi et al., 2014).

1.3.3 Cell size regulatory pathways and molecules

Numerous pathways have been described that regulate cell size. Here I give a short overview of three selected examples: two fundamental, cell size related pathways and an independent cell size regulator. As I have shown above a common example for the fission yeast and budding yeast from the TOR pathway, I start with TOR pathway.

1.3.3.1 TOR pathway

In 1991, two proteins were identified in budding yeast as they can be blocked by a macrolide, called rapamycin that causes immunosuppression in humans. These proteins were named **Target of Rapamycin**: **TOR1** and **TOR2** (Heitman et al., 1991). Later it was shown that these TOR proteins are involved in growth control (Kunz et al., 1993). Studies in yeasts have identified that these proteins form two distinct protein complexes (TORC1 and TORC2) (see Figure 1.3-2) and fulfil crucial roles in connecting metabolism with cell size regulation (Martin et al., 2004; Wullschleger et al., 2006). It is now known that TORs are evolutionally conserved serine/threonine kinases that are controlled by environmental conditions. TORs are in a signalling network and phosphorylate several proteins (see Figure 1.3-3) (Gonzalez & Rallis, 2017).

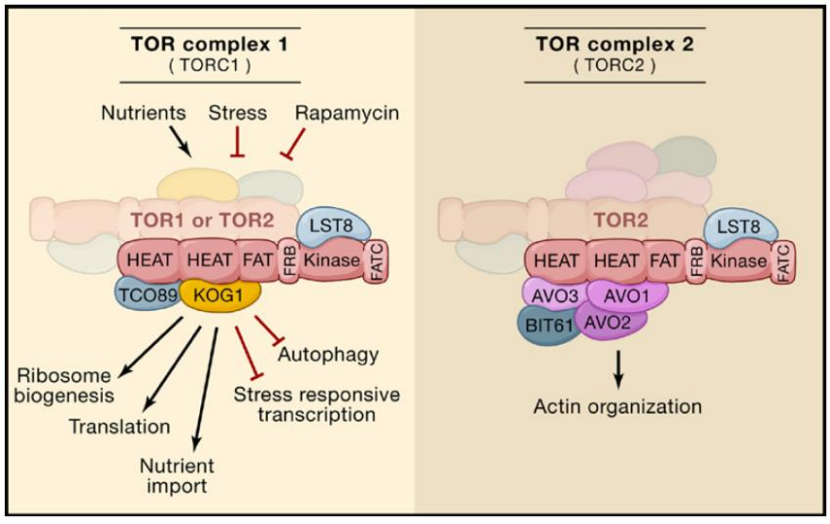


Figure 1.3-2: TORC1 and TORC2 complexes in budding yeast.

Diverese effects of TORC1 and TORC2 in buddying yeast.

The figure is reproduced from Wullschleger et al. (2006).

TORCs have various signal inputs from stress to nutrient availability and regulate multiple processes (Wullschleger et al., 2006). The first complex, TORC1 is responsible for promoting ribosome biogenesis, protein anabolism and cell proliferation, while the second complex TORC2 has antagonistic effects (Averous & Proud, 2006). The rate of ribosome biogenesis is managed by nutrient conditions and transmitted by the TOR pathway also *via* Sch9 proteins (Jorgensen & Tyers, 2004). TOR pathway has also been identified in mammalian species (named mTOR) (Brown et al., 1994). In mammals, only one mTOR protein exists compared with two protein in yeasts (Hay & Sonenberg, 2004; Wullschleger et al., 2006). For a comparison between processes in metazoans and yeast see Figure 1.3-3.

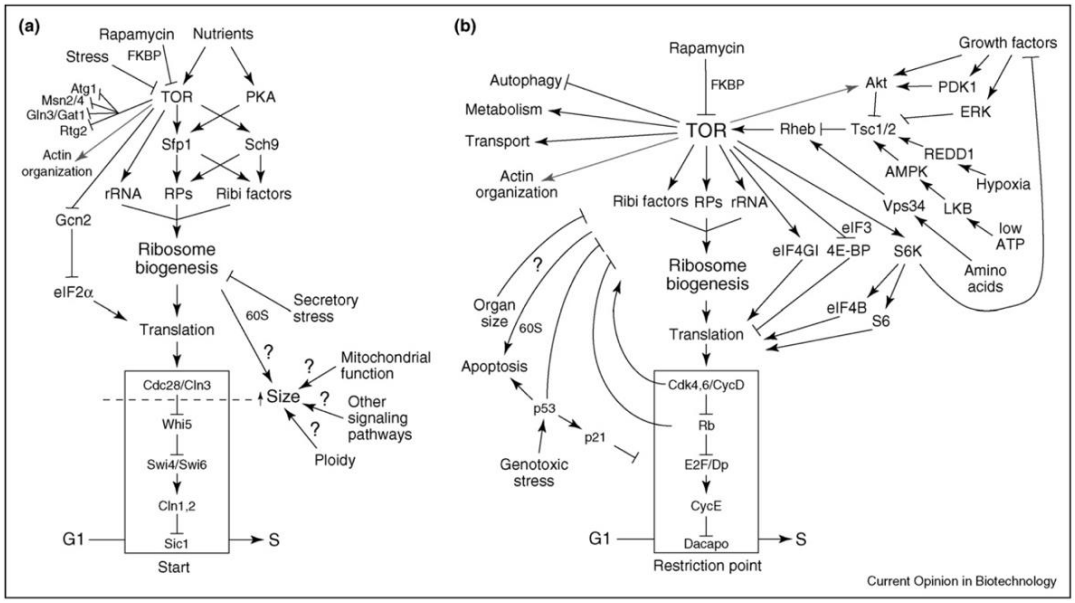


Figure 1.3-3 Representation of the major pathways regulating growth and division at the G_1 to S phase transition in budding yeast and Drosophila melanogaster S2 cells.

(A) In budding yeast, the G_1/S transition and the crucial role of TOR.

(B) G_1/S transition in metazoans, similar mechanisms exist as for budding yeast, as described in the text.

The figure is reproduced from Cook & Tyers, 2007.

1.3.3.2 Hippo pathway

The Hippo pathway was discovered in studies on *Drosophila Melanogaster*. Subsequently, it was shown to be a highly conserved pathway responsible for organ size and tissue homeostasis (Yu et al., 2015). The Hippo pathway regulates cell proliferation and apoptosis (Yu et al., 2015). The mechanism involves Hippo pathway members inhibiting YAP/TAZ to facilitate gene transcription (Figure 1.3-4). Transcription is regulated in part by phosphorylation of the transcriptional regulator yes-associated protein (YAP) (Huang et al., 2005), YAP is under the direct control of LATS1/2 (Zhao et al., 2010). YAP controls transcription of *CCNE,* which encodes the cell cycle regulator Cyclin E and the gene encoding the cell death inhibitor *DIAP1* (Huang et al., 2005; Nolo et al., 2006; Tapon et al., 2002; Thompson & Cohen, 2006). In mammalians, CTGF (connective tissue growth factor) pathway is a usual marker of the YAP activation (Zhao et al., 2008). Members of the Hippo pathway are directly affected by the modifications of the actin cytoskeleton, for *e.g.* by the overexpression of Rho GTPases (Dupont et al., 2011). Loss of function mutations in *Drosophila* such as Capulet gene directly leads to tissue overgrowth (Sansores-Garcia et al., 2011).

The best-known function of Hippo pathway is to regulate tissue and organ homeostasis. It has been shown in *Drosophila* if Hippo pathway regulators such as Kibra, Mer and Ex or Hippo pathway kinases such as Wts and Hpo are upregulated, this leads to an enlargement of organs and overgrowth of wings and eyes (Halder & Johnson, 2011; Pan, 2010; Yu et al., 2015). Hippo pathway members affect the apico-basal polarity of cells, that regulates cell growth (Genevet & Tapon, 2011). Mutations of the members of the Hippo pathway can lead tumour development and cell transformation in mouse liver (Dong et al., 2007). It seems that Hippo pathway plays a fundamental role in organ size control in model organisms, while members of the pathway play crucial roles in the development tumours (Harvey et al., 2013).

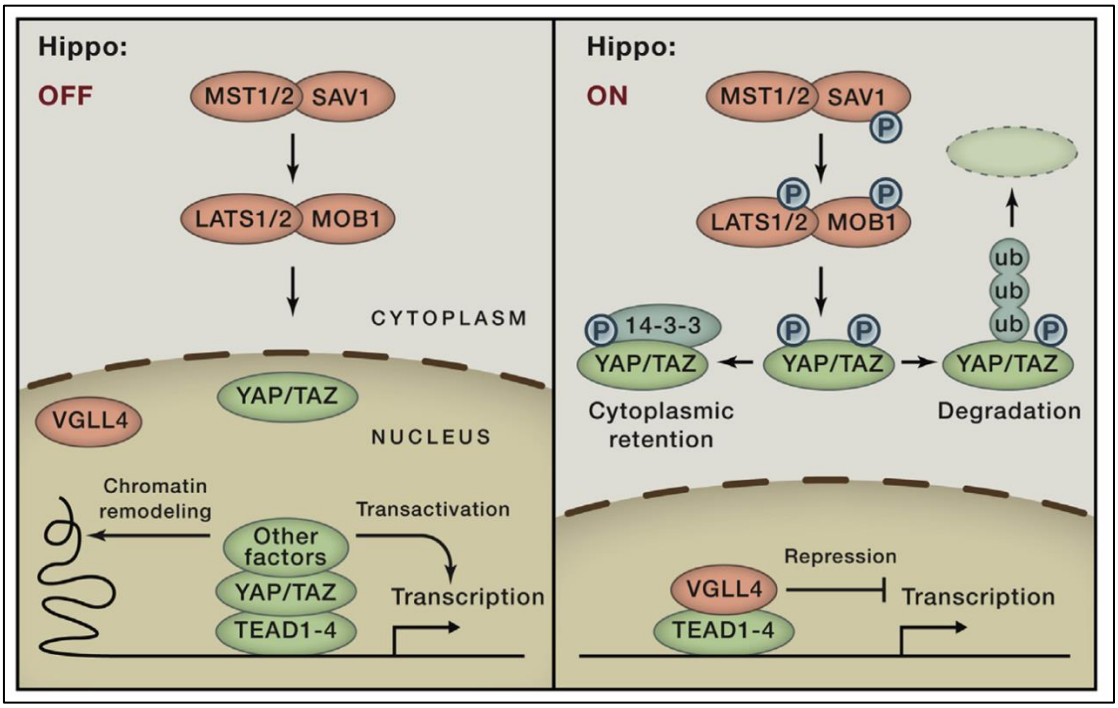


Figure 1.3-4: Inhibition of YAP/TAZ Transcriptional Coactivators by LATS1/2

(Left) In this case, the Hippo pathway is off. The key molecule is YAP/TAZ, which enters the nucleus, then competes with VGLL4 for TEADs, and recruits other factors to facilitate gene transcription. **(Right)** In this case, the Hippo pathway is on. The YAP/TAZ complex is phosphorylated by LATS1/2 on multiple sites, resulting in interaction with 14-3-3 and cytoplasmic retention; phosphorylation also leads to YAP/TAZ poly-ubiquitination and degradation. VGLL4 interacts with TEADs and represses target gene transcription.

The figure is reproduced from Yu et al. (2015).

1.3.3.3 Largen protein

In 2014 Yamamoto *et al.* found in human cells that the product of the *PRR16* gene is Largen, which is a protein that regulates cell size in *in vitro* experiments (Yamamoto et al., 2014). Largen reportedly initiates the translation of specific sub-sets of mRNAs, that encode mitochondrial related protein functions (see Figure 1.3-5) (Yamamoto et al., 2014). *PRR16*$^{-/-}$ transgenic mice have increased cell size and accelerated mRNA translation and mitochondrial function (Yamamoto et al., 2014). The Largen protein is an example of the fact that there are independent cell size regulators from main signalling size and organ control pathways (Yamamoto & Mak, 2017).

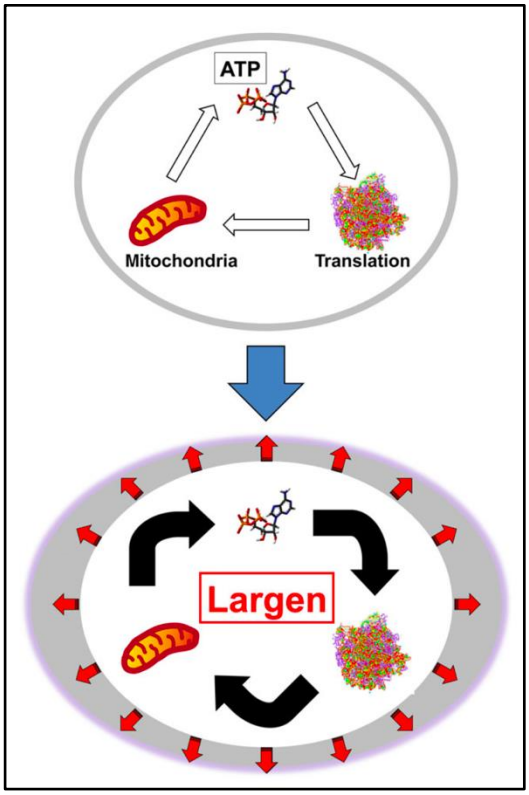


Figure 1.3-5: A proposed model for the cell size control by Largen

Mitochondria produce ATP, which is used for many cellular processes, such as mRNA translation. When Largen is overexpressed, cells will have more mitochondria and more mitochondria lead to the production of more Largen in a positive feedback loop.

The figure is reproduced from Yamamoto & Mak (2017).

1.3.3.4 Cell size regulation in my thesis

In this chapter I have reviewed evidences that the regulation of cell size is a highly complex process, which involves many control molecules (Lloyd, 2013). Some proteins actively regulate cell size, such as MAPK dependent protein p38 (Liu et al., 2018) or TOR pathway member proteins (Gonzalez & Rallis, 2017). Other proteins have to be present for cells to increase in size and the effects of a knock-out or knock-down on cell size are due to secondary effects of other processes such as cell proliferation or the production of mitochondria (Meziane et al., 2011).

In my thesis, I'm will investigate cell size regulation as a process, where a loss of the function of an individual protein can alter the overall size. Some of these mutants might be associated with perturbations in growth (e.g. defects in ribosomal proteins), while others might be true regulators of cell size. This second group is the main target of my investigations. I am aware that in multicellular organisms, tissue and body level compensation can affect cell size, thus I will perform cellular level analyses of cells in tissues of the plant *Arabidopsis thaliana* and individual cells in culture, using T cells isolated from human peripheral blood.

1.4 Orthology in Biology

The evolution of life on Earth started about 3.5 billion years ago based on the fossil record (Schopf et al., 2002). Some genes that occur in different species are highly conserved throughout the phylogenetic tree, while others are transient or specific to particular species (Fang et al., 2010). Each organism has a set of genes that lead to phenotype. These traits may be the same in different organisms, while some can be extremely different. For example, the *CDK1* gene, which encodes Cdc2/Cdk1 is an evolutionarily conserved gene that regulates mitosis from yeast to humans (Malumbres & Barbacid, 2009). By contrast, the IFT proteins are components of cilia motor proteins that regulate motility of lower organisms, but these evolutionarily conserved proteins regulate T-cell receptor internalisation and vesicular trafficking in human T-lymphocytes (Finetti et al., 2009; Onnis et al., 2016). A 2010 study named the genotype-phenotype association of gene homologues in different species "*phenologs*" – named as phenotypes related by the orthology of the associated genes in two different organisms (McGary et al., 2010). A particular gene deletion or duplication observed in various organisms was not evolutionarily conserved by accident. Instead there is a specific reason driven by the associated phenotype that has a selection benefit (Fang et al., 2010; Fitch, 2000). Gene duplication events are also key accelerators of novel functional gain (Magadum et al., 2013). However, new gene duplication is more likely to end as a pseudogene then a gene with a novel function, unless there is a positive selection pressure (Walsh, 1995).

1.4.1 Conserved Genes

Gene duplication and loss of function all generate genetic diversity and are processes that are under selection pressure. Evolutionary pressure on alternative splicing can create "hotspots" within a protein sequence, that are evidently selected for a specific function (Xing & Lee, 2005). A study by Pagani *et al.* (Pagani et al., 2005) showed that introduction of synonymous substitutions into exon 12 of *CFTR* led to exon skipping and an inactive protein for approximately one third of the substitutions tested, which has been proposed to be an RNA-level selection pressure (Xing & Lee, 2005). On the other hand, some genes are highly conserved in many organisms, which means they have kept the same genetic function and sequence similarity in different species. Examples are 18S and 28S ribosomal RNAs and the protein binding domains of ABC transporters (Isenbarger et al., 2008). Experimental studies showed that the primary sequences of 16S ribosomal RNA in prokaryotes and 12S ribosomal RNA in the mitochondria of eukaryotes are highly conserved (Zwieb et al., 1981). Moreover, the eukaryotic 18S rRNA shows only 0.1% sequence divergence between human and mouse (Gonzalez & Schmickel, 1986). These rRNAs are involved in a key and essential function of cells, *i.e.* translation that needs to be maintained for cells to survive and out-compete others in the population (Isenbarger et al., 2008). There is a consensus in the literature that the conserved genes usually have high sequence similarity (>90%) and few duplication events, although detailed sequence analysis shows smaller sequence similarity, and higher sequence variability for some genes (Martinez-Porchas et al., 2017). In many cases, genes encoding proteins are transient or very specific to some functions in some species in which deletion, neofunctionalization and/or non-orthologous gene displacement (Koonin et al., 1996) could end up having novel functions (Fang et al., 2010).

1.4.2 Terms used in Orthology

The high functional similarity of proteins in different species or homologous functional and sequence relationships in the same species can be categorised into two classes (Figure 1.4-1):

(1) *Orthologs*: are homologous genes in different species that evolved from a common ancestral gene by speciation (radiation). Some orthologs retain the same function during the course of evolution (Fang et al., 2010). Walter Fitch first described the term Ortholog in 1970 (Fitch, 1970).

(2) *Paralogs* are homologous genes related by duplication within a genome and usually evolve new functions (Cotton, 2005). The latter definition can be divided into two subcategories, depending on the timeline of the speciation event (Heidelberg et al., 2002):

a. *outparalogs*, if the duplication happened before speciation
b. *inparalogs*, if the duplication event occurred after speciation.

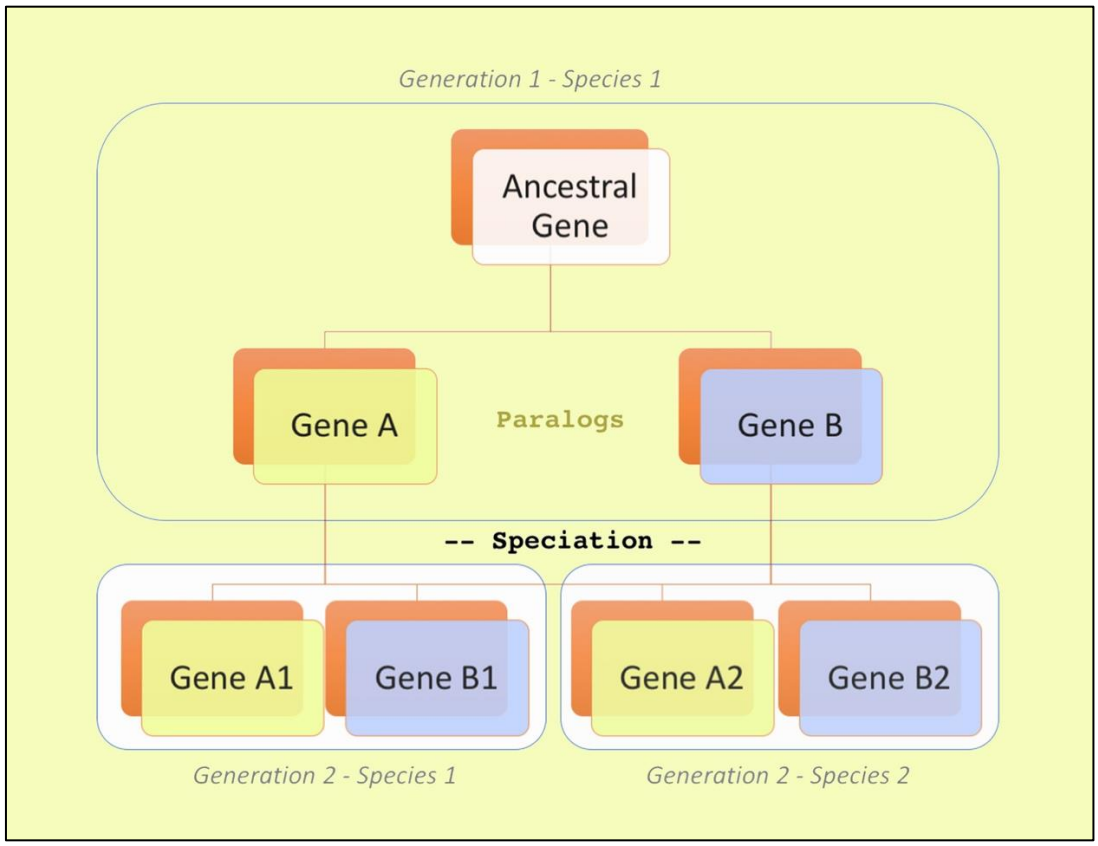


Figure 1.4-1: Orthologs and paralogs

Orthologs are homologous genes that were created from an ancestral gene with a speciation event, while paralogs are homologous genes that were created with a duplication event. **Gene A** & **Gene B** are paralogs of each other because they were created by a duplication event. **Gene A1** & **Gene B1**, as well as Gene A2 & Gene B2, are the same relation, while **Gene A1** & **Gene A2** as well as **Gene B1** & **Gene B2** are orthologs of each other.

The figure is based on Jensen (2001).

1.4.3 Orthology Databases

Several published orthology databases exist, which are specialised at least into four different aspects (Jensen et al., 2008):

(1) the first type cluster pairs of genes with the same biological functions (Jensen et al., 2008; O'Brien et al., 2005). These databases use a Bi-directional Best Hit (BBH) strategy, which simply means that these are the reciprocally most similar pair of genes in two species. A notable example of this category is InParanoid (O'Brien et al., 2005; Ostlund et al., 2010).

(2) The second type is clustering with more species using the best hit method. The best example of this is type is orthoMCL database (Li et al., 2003).

(3) The third type uses a clustering method of three different genomes, called clustered triangles of mutually consistent, genome-specific best hit genes *e.g.* the COG/KOG database (Tatusov et al., 1997).

(4) While finally, the fourth type uses phylogenetic trees to identify functional divergence events, notably TreeFam (Li et al., 2006; Thomas et al., 2003; van der Heijden et al., 2007). Only one of these contain functionally annotated groups of genes, namely the COG/KOG database (Tatusov et al., 1997).

These database concepts employ different mechanisms to identify the best possible orthologs in different organism. In my thesis, I investigated several orthology databases, then finally I used the eggNOG database (Huerta-Cepas et al., 2018; Jensen et al., 2008; Muller et al., 2010; Powell et al., 2012, 2014) to connect genes and proteins throughout the phylogenetic tree to address my overarching aim of identifying novel cell size regulators.

1.4.4 Using Orthology Databases

In my thesis I use several Orthology databases. Specific databases identify orthology either by genomic similarities or by functional similarities of proteins in different organisms.

As of March 2018, according to the research curation website *Quest for Orthologs* (https://questfororthologs.org/orthology_databases) there are more than 30 published ortholog databases which are publicly available. During my research I tested most of them, but in the early phases of my PhD I used the five databases which contain the seven species of interest described in Chapter 2. The five databases are:

(1) **KOG/COG** (http://www.ncbi.nlm.nih.gov/COG/) (Tatusov et al., 1997). Published in 1997 as the first database to use clusters of orthologs (COG) approach, based on their functional similarity. This database has a major advantage of introducing a novel "COG" concept, however it has not been updated since 2003 (Tatusov et al., 2003).

(2) **eggNOG** (http://eggnog.embl.de) (Huerta-Cepas et al., 2016; Jensen et al., 2008), which uses a similar approach to find functional novel COG/KOG groups. eggNOG contains several subsets of orthologue relations between realms of the phylogenetic tree, which are collected in functional orthologue groups. I selected this database to use for the analyses described in my Thesis for the reasons described in Chapter 2.2.4.1 .

(3) **HomoloGene** (http://www.ncbi.nlm.nih.gov/homologene) (Coordinators, 2014). An NCBI based project that aims to find homologous proteins based on their sequence similarity. Closely related proteins are first matched, then they are placed on a phylogenetic tree that is expanded with other proteins, based on their genetic distance. The major disadvantage is of HomoloGene is that it focuses on genetic similarity rather than functional orthology.

(4) **inParanoid** (http://inparanoid.sbc.su.se/) (O'Brien et al., 2005; Ostlund et al., 2010). An algorithm Is used to make a pairwise comparison of genes in two species and then creates orthologue groups. This database contains groups of orthologous genes, called paralogs that were created by a speciation event. This database is recommended if the aim is to identify pairwise orthologs. However, it Is not able to identify groups of ortholog. I used it in the early phase of my research but chose to use the eggNOG database instead (see (2) above).

(5) **orthoMCL** (http://orthomcl.org/orthomcl/) (Li et al., 2003). This database clusters and groups orthologically related proteins by an all-against-all BLAST method and is based on sequence similarity. Unlike inParanoid, orthologous groups in this database are not limited to a pairwise comparison of two species. OrthoMCL has for the advantage of Identifying groups of orthologs, however it does not predict functional orthologs.

1.4.4.1 Selected Orthology Database

As my concept for orthological analysis was to find functional orthologs in other organisms that regulate cell size I decided to use a database that provides the best functional ortholog proteins in other organisms. The cell size attribute is a quantitative trait and if one protein in an ortholog group is identified as having a role in regulating cell size, the same attribute may also hold for other members of this ortholog group. After careful analyses of other databases, I decided to use eggNOG. This database is regularly updated and contains good quality and grouped results. Moreover, it provides scripts that makes it easier to handle the data.

1.4.4.2 eggNOG database

The eggNOG database (abbreviation of the **e**volutionary **g**enealogy of **g**enes: **N**on-supervised **O**rthologous **G**roups) contains orthologous groups that were created by extending the COG algorithm with a graph-based unsupervised clustering method to create a genome-wide overview of orthology (Jensen et al., 2008). These clusters were derived by identifying common markers for the genes based on their textual descriptions, and annotated functional categories (Jensen et al., 2008). To create orthologue groups, sequences were extracted by Smith-Waterman alignments (Smith & Waterman, 1981), which identified the best reciprocal matches in *all-against-all* pairwise comparison. This alignment comparison data was provided by the SIMAP project (Kuhn et al., 2014; Powell et al., 2012). Taxonomy levels were predefined. The eggNOG algorithm took the basis of Clusters of Orthologous Groups (COG) in three realms of the phylogenetic tree: COGs (for Bacteria) (Galperin et al., 2015), KOGs (for Eukaryotes) (Tatusov et al., 2003), and arKOGs (for Archea) (Makarova et al., 2015). Every updated orthology group is benchmarked and compared with similar approaches in the past, such as OrthoBench (Trachana et al., 2011, 2014). Annotations were hierarchically consistent in nested groups. Functional annotations were made *via* KEGG pathways (Kanehisa et al., 2014), SMART/Pfam protein domains (Finn et al., 2014, 2016), Gene Ontology (Ashburner et al., 2000; Gene Ontology Consortium, 2015) and COG functional categories (Tatusov et al., 2003). Finally, the eggNOG database contained hierarchically consistent, functionally relevant orthologue groups for 2031 eukaryotic and prokaryotic organism (as of eggNOG latest version 4.5, 2016) (Huerta-Cepas et al., 2016).

Chapter 2

Bioinformatics analysis of cell size regulation conservation

2.1 Introduction

As I have described in the introduction, the regulation of cell size is a fundamental but barely understood attribute of biological cells (Cook & Tyers, 2007). While a researcher can clearly see that certain types of cells are likely to maintain the same cellular dimensions, it is critical to understand how cells regulate their size. Some genes are conserved throughout the phylogenetic tree spanning millions of years of evolution (Fujiwara et al., 2002). In some cases, genes are maintained during the separation of ancestor species to different daughter species such as in the case of *S. cerevisiae* and *S. pombe,* that happened 350 million years ago (Hoffman et al., 2015). For plants and animals, this happened much earlier, about a billion years ago (Knoll, 1992; Sipiczki, 1995). However, it can be seen that some gene sequences are more similar in *S. pombe* and *H. sapiens* than in *S. pombe* and *S. cerevisiae*, from phylogenetic trees based on rRNA (Sipiczki, 2000). In addition, common pathways such as the TOR pathway are conserved and affect the size of the cells in distant organisms in a similar way (Edgar, 2006). This is also supported by the notion that Protein-Protein Interactions (PPIs) between proteins belonging to a common protein complex are more conserved than PPIs between proteins from different complexes (Beltrao et al., 2012).

2.1.1 Aims of my Bioinformatics studies

In this chapter, therefore I address the question whether there are evolutionarily conserved genes encoding proteins which affect cell size. I use information about seven different species, some separated by a billion years of evolution as well those separated by just a few hundred million years. The species are *Arabidopsis thaliana, Caenorhabditis elegans, Danio rerio, Drosophila melanogaster, Homo sapiens, Saccharomyces cerevisiae* and *Schizosaccharomyces pombe.* I will explain the details of the data obtained from each organism below in this chapter. My aim is to identify which proteins hold conserved functions of cell size regulation. I will determine, whether by analysing conserved orthologous groups and dropping common growth regulators also affecting cell size I can propose potential cell size regulators in various organisms.

2.1.2 Bioinformatics as a field of science

Bioinformatics as a field can be described as "*the application of computational tools to organize, analyze, understand, visualize and store information associated with biological macromolecules*" (Diniz & Canduri, 2017; Luscombe et al., 2001; Pevsner, 2015). Another definition is "*The science of collecting and analysing complex biological data such as genetic codes*" (Oxford Dictionary). Bioinformatics is a large and quickly expanding field of biology. Novel approaches, methods and databases are regularly published by bioinformatics researchers who can address existing biological questions from a novel viewpoint.

2.1.2.1 Origins of Bioinformatics

Origins of the bioinformatics field date back as far as the 1960s when DNA sequencing became feasible (Hagen, 2000). Large computers became common at universities in the USA and researchers started to use them to solve biological questions (Hagen, 2000). With the help of IBM's first high-level programming language, called FORTRAN (formula translation), researchers were able to determine protein sequences with coded programs (Dayhoff & Ledley, 1962). The first major step in the direction of computerized data collection was made by Margaret O. Dayhoff, who published a book on collected protein data called *Atlas of Protein Sequence and Structure* first printed in 1966 (Eck & Dayhoff, 1966). This book was cited more than 4500 times, while the collection of data evolved to a *Protein Information Resource Database*, which was the first online database that was publicly available (Hunt, 1983). Based on these data several researchers developed early program codes to analyse phylogenetic evolution based on cytochrome c trees (Dayhoff, 1969) and sequence searches (Doolittle, 1997). By the 1970s many of the techniques had developed improve the task of analysing proteins. Novel methods also emerged to study DNA nucleic acid sequences (Diniz & Canduri, 2017).

2.1.2.2 Biological databases

As biological research has evolved throughout the years, bioinformatics databases have appeared to store and organise large amounts of data. These databases can store, handle, search, share and sometimes analyse the data from a wider perspective than can be obtained from biological experiments on a single species. Databases can be categorized based on their specificity to organisms, level of molecular data, function and usage of source information. Specificity can be to a certain organism, a certain group of organisms or all organisms, while the level of molecular data can be at DNA, RNA or protein. The function of a given database could be just a simple collection of experimental data or it could provide a relation to existing data points by some functions, such as functional pathways or homology. As of January 2018 according to the journal *Nucleic Acids Research* there are 1,738 publicly available biological related databases (the list of databases can be found on http://www.oxfordjournals.org/nar/database/c/) (Rigden & Fernández, 2018).

2.1.2.2.1 Organism-specific databases

Organism-specific databases can be found for all highly investigated model species in bioinformatics. For those species I focused on in my research, there was at least one species-specific database available. These databases usually summarize all the available experimental data and information for all genes and proteins inside the given species.

(1) *A. thaliana:* TAIR database (https://www.arabidopsis.org/) (Huala et al., 2001)

(2) *C. elegans:* WormBase (http://www.wormbase.org/) (Van Auken et al., 2012)

(3) *D. rerio:* ZFIN database (https://zfin.org/) (Bradford et al., 2011)

(4) D. melanogaster: FlyBase database (http://flybase.org/) (St Pierre et al., 2014)

(5) *H. sapiens:* There are many specific databases. In my research, I investigated closely the MitoCheck / Cellular Phenotype Database (https://www.ebi.ac.uk/fg/sym/study/M1_SyM) (Hutchins et al., 2010;

Neumann et al., 2010), a database containing the genome-wide phenotypic profiling of each of the human protein-coding genes.

(6) *S. cerevisiae:* Saccharomyces Genome Database (https://www.yeastgenome.org/) (Hirschman et al., 2006)

(7) *S. pombe:* PomBase database (http://www.pombase.org/) (Wood et al., 2012)

2.1.2.3 Predicting function using homology

As I pointed out earlier in the Introduction (Section 1.4.1), genes can be conserved in many species and they can share high sequence and functional similarity (Isenbarger et al., 2008). The biological concept of finding common ancestors of genes has been with us for a long time, since the birth of bioinformatics (Pearson, 2013). Homologous proteins share similar molecular structures and frequently they share similar functions. The main issue is that sequence similarity is more predictable and countable, while functional similarity is less predictable. The most difficult part of the prediction is that one given protein can have multiple, sometimes different functions (Loewenstein et al., 2009).

Bioinformatic methods have evolved to find similar structures in different proteins (Pearson, 2013). Scripts have been written to find similar DNA or protein sequences in different organisms by pairwise or all-against-all strategies. Commonly used similarity search programs which search for sequence similarities of proteins include BLAST (Altschul et al., 1997), SSEARCH (Smith & Waterman, 1981), FASTA (Pearson & Lipman, 1988) and HMMER3 (Johnson et al., 2010). The main aim of these programs is to infer relationships between DNA sequences and to predict homologous protein pairs more accurately than if they were selected by chance (Pearson, 2013). The results obtained are mainly sequence and structural domain homologs.

Methods have also evolved to find multiple sequence alignments in protein families such as CDD (Marchler-Bauer et al., 2007) or PROSITE (Hulo et al., 2008) or a different profile-hidden Markov model such as Pfam (Finn et al., 2008). As homologous proteins

evolve, they form distinct 3D structures. Bioinformatic approaches have also arisen to find structure-based family signatures that classify their protein domains into families, such as CATH (Orengo et al., 1997) and SCOP (Murzin et al., 1995). The reasoning behind using these protein family models are that common domains in proteins may share the same function. However, predicting accurate functional orthologs is very complicated and it is mainly based on sequence or structural element predictions. I needed to use a protein functional ortholog resource and so I used the eggNOG database (Jensen et al., 2008), which contains functional ortholog protein groups from thousands of species, including the seven species of interest for my thesis.

2.1.2.4 Gene Ontology (GO)

Gene Ontology (GO) (Ashburner et al., 2000) database is a well-defined functional, bioinformatics database whose primary goal is to represent all genes and gene product by all related biological attribute in a structured way. The GO database can be reached either via APIs or via a web-based AmiGO2 tool (http://amigo.geneontology.org/) (Carbon et al., 2009). GO is a part of Open Biomedical Ontologies (OBO), a research wide biomedical effort to collect controlled and checked vocabularies for distinct fields of biology and medicine. Gene Ontology started as a major bioinformatics initiative in 2000 that addresses the following three main goals:

(1) represent an up-to-date vocabulary for genes and gene products properties
(2) annotate genes into broader and more-specific ontologies
(3) provide a widely available, structured bioinformatics tool for gene annotation and annotation enrichment analysis for researchers

2.1.2.4.1 GO terms

Gene Ontology terms are subdivided into three ontology categories that represent three distinct areas of biology as below:

(1) **cellular component**: ontology describes the location of the cellular component at the subcellular location and at the macromolecular complex level. This localization also includes multi-subunit enzymes and other protein complexes.

(2) **biological process:** ontology cast a series of events or molecular functions into a specific biological process. The criterion for being a process: it must have a defined beginning and an end. Processes have hierarchical and parallel relation to each other. A given process can be for example a part of, a child of or a parent of another biological process annotation.

(3) **molecular function**: ontology describes the distinct abilities that related to the annotated component. Terms can be linked to various specific activities that include transporting, binding, holding and transforming molecules.

2.1.2.4.2 GO Annotations

An annotation in the GO database always holds an "evidence code" that explains how that annotation came to light. Three types of evidence codes are used in GO to refer to the association of the term and the biological product (reviewed in (The Gene Ontology Consortium, 2017). I only used the first type in my experiments.

(1) *experimentally-supported annotations (EXP):* These updates made by PhD-level experts, who curate peer-reviewed scientific literature. To maintain the quality and consistency of the database curators met regularly and discussed further developments, as well as they do annotation consistency exercises. The Gene Ontology database can be updated from manually curated, experimental based sources, such as from microRNA database (Huntley et al., 2016) as well as from protein-protein interaction (PPI) databases such as IntAct (Meldal et al., 2015).

(2) *Phylogenetically-inferred annotations (IBA):* There are specific phylogenetic approaches to expand annotations based on phylogenetic. Curators have developed software like **PAINT** (Phylogenetic Annotation Inference Tool) (Gaudet et al., 2011). The curators of the GO can model of gain and loss of function in genes at distinct branches in a given phylogenetic tree. Curators use this to infer novel annotation for gene families (The Gene Ontology Consortium, 2017).

(3) *Computationally-inferred annotations (IEA):* These updates are made by software, and associated terms are not individually reviewed. There are three methods taking place to add computationally inferred annotations to the GO database: (1) comprehensive methods based on sequence homology, and group of homologous protein using InterPro2GO tool (Mitchell et al., 2015), a similar initiative to my approach; (2) Based on UniProt controlled vocabulary terms, that describes an enzymatic activity; (3) Then based on orthologs inferred from Ensembl gene trees, an automatic approach to

transfer novel annotations based on one gene 1:1 orthologs in other species in the same taxonomic clade.

2.1.2.4.3 Process of curation

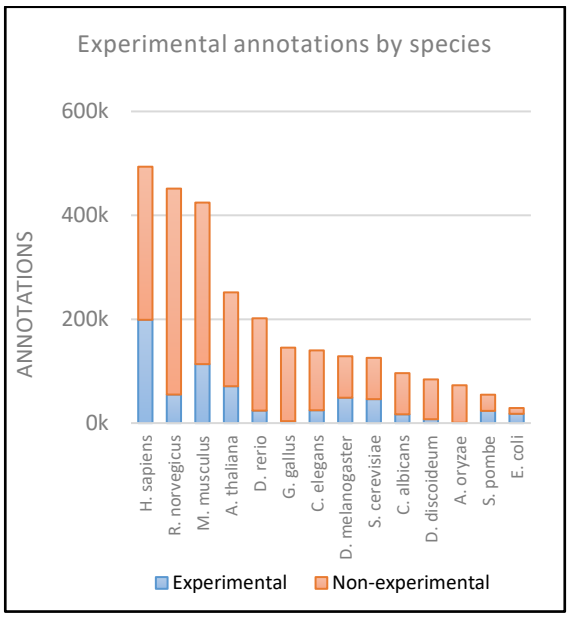
Gene Ontology Consortium (GOC) maintains and updates the database (Ashburner et al., 2000). Annotation of terms in the GO database can be either manually curated, based on published literature or computationally generated (Camon, 2004). Specific expansion of GO can be made via cross-discipline development projects that aim an expansion of a specific research field such as immune system, heart development, transcription and cell cycle (Dessimoz & Walker, 2017). GO can be updated by community contributors of the GO database. The majority of these contributors are specific research groups, who focuses on a given model organism, these are such as **ZFIN** (Bradford et al., 2011), **PomBase** (McDowall et al., 2015) and **FlyBase** (Tweedie et al., 2009), or on a broader view at the proteomic level like the **UniProt** database community (Huntley et al., 2015). Independent researchers can also update the GO by contacting the GOC and they forward the request to the appropriate research community. The current list of scientific communities that contributes to the GO database can be found on http://geneontology.org/page/go-consortium-contributors-list.

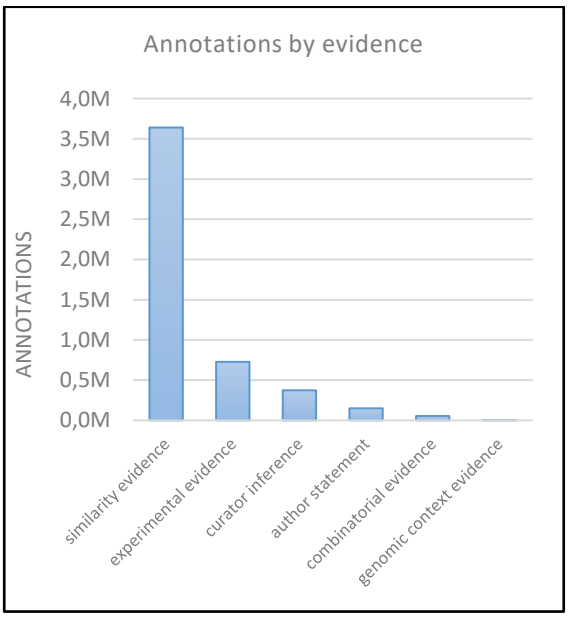
2.1.2.4.4 Quality control on the curation

The GO community runs distinct quality controls to maintain the biological relevance and quality of the database (Poux & Gaudet, 2017). GO usually runs automated checks via GO annotation rulebase (http://geneontology.org/page/annotation-quality-control-checks), that validates the relation between the biological content and syntactic in sources. In the other hand, GO consortium regularly ask researchers to independently annotate the same original paper to ensure the annotation is of high quality (Poux & Gaudet, 2017). As a third way, there are bioinformatics tools like the Reference Genome Project that provide a useful tool to improve annotations in line with GO (Gaudet et al., 2011).

2.1.2.4.5 Statistics on GO database

As of March 2018, GO contains **4,945,480** annotations (latest statistics retrieved from: http://amigo.geneontology.org/amigo/base_statistics). Most of these annotations are given based on similarity, not by direct experiments in each organism (Figure 2.1-1). Data on more and more tested organisms contain experimental evidence. Although annotations of less investigated organisms are mostly predicted and based on orthology. In the case of data on human cells, there are 294,776 non-experimental and 198,810 experimental annotations.





Annotation name	Number
similarity evidence	3,640,710
experimental evidence	727,191
curator inference	373,362
author statement	149,552
combinatorial evidence	53,578
genomic context evidence	1,087

Figure 2.1-1: Basic statistics of Gene Ontology

The left chart describes the proportion of experimental and non-experimental GO annotation in distinct species. The chart on the right gives the evidence for these annotations. The non-experimental, similarity-based evidence are the most common.

2.2 Materials and Methods

2.2.1 Creating a comprehensive database of orthologous proteins

Data on the following seven, evolutionarily distant eukaryotic species were obtained during this research: *Arabidopsis thaliana, Caenorhabditis elegans, Danio rerio, Drosophila melanogaster, Homo sapiens, Saccharomyces cerevisiae* and *Schizosaccharomyces pombe.* I was collecting the list of all proteins involved in cell size regulation in these species.

2.2.2 Sources for existing cell size regulators

2.2.2.1 S. cerevisiae

I used the study by Jorgensen *et al.* (2002) (Jorgensen et al., 2002) as the source of known cell size regulator proteins in *S. cerevisiae*. In this study, the authors used the complete budding yeast knockout gene deletion collection of the 4,812 open reading frames (ORF) of individual genes and investigated their effects on cell size. Clusters for cell size distribution were established: 451 abnormally small (*whi*) or large (*lge*) mutants were found in 4 groups (haploid or diploid, large or small mutant). I added to this list the cell size regulators from the study by Moretto *et al.* (2013) (Moretto et al., 2013), which expanded the previous results by Jorgensen *et al.* (2002).

2.2.2.2 S. pombe

To collect *S. pombe* cell size regulator genes, an academic study performed by Hayles *et al*. (2013) was used (Hayles et al., 2013). This study focuses on creating a genome-wide gene deletion library of 4,843 fission yeast genes, while the authors investigated the effect of each deletion on cell shapes (see Table 2.2-1), I used only the size or shape specific attributed gene deletion candidates. I selected the terms that might represent changes in size based on a discussion with the first author of the original study. Plus I added extra 90 genes that has been described to affect cell size distribution (Graml et al., 2014). I will refer to these misshapen cells commonly as "*size mutants*" in my thesis.

Table 2.2-1: Usage of gene attributes in *S. pombe* based on Hayles et al., 2013

Shape form	Usage in this analysis
curved	***Used***
germination	Not Used
long branched	***Used***
long high penetrance	***Used***
long low penetrance	***Used***
misshapen essential	Not Used
misshapen viable	Not Used
misshapen weak viable	Not Used
rounded	***Used***
skittle	***Used***
small	***Used***
spores	Not Used
stubby	***Used***

2.2.2.3 H. sapiens

To collect *H. sapiens* cell size related genes, the **MitoCheck** database (http://www.mitocheck.org) was used, which is based on the paper published in 2010 (Neumann et al., 2010). The MitoCheck database contains the genome-wide phenotypic profiling of each of the human protein-coding genes in the human genome database. Each gene was uniquely targeted with a short interfering RNA (siRNA) library, and the effect on cell cycle progression was monitored by an assay using cells expressing a fluorescent core histone protein (Hutchins et al., 2010). Each of the cell lines was analysed by two-day live cell imaging of fluorescently labelled chromosomes, which were annotated according to 16 morphological classes (Neumann et al., 2010).

MitoCheck contains information about cell size phenotypes and to build the database I used the "*Large"* phenotype class of the primary screen results from the MitoCheck database (small cell or small general classification were not categorized in MitoCheck primary screen phenotypes).

2.2.2.4 D. melanogaster

To collect *D. melanogaster* cell size related genes, a study published by Bjorklund *et al.* (2006) (Björklund et al., 2006) was used. These authors analysed the effect of the loss of function of 70% of Drosophila genes (including 90% of genes conserved in humans) (Björklund et al., 2006). Several properties were measured including the effect on cell size. I have focused on all of the genes whose loss of function affected cell size and merged this data with info from the Drosophila specific **FlyBase** database (St Pierre et al., 2014) on cell size defective phenotype.

2.2.2.5 A. thaliana, C. elegans and D. rerio

For three species, namely *A. thaliana*, *C. elegans* and *D. rerio* there are no reports describing genome-wide mutational analyses of cell size and growth. Therefore I used a biological process annotation term from Gene Ontology (GO) (Ashburner et al., 2000), called GO:0008361 "*regulation of cell size*" to identify cell size regulators.

This same term contained several genes I have found for other organisms, but the original data from those genome-wide screens included many more genes than I found by the same approach in the Gene Ontology database.

2.2.3 Incorporation of protein-protein interaction (PPI) data

Since conserved biological functions are mostly pathway dependent, I hypothesised that proteins that interact with multiple cell size regulator proteins in a given organism might also be involved in cell size regulation or their orthologs could play similar roles. Thus, I extended the database of cell size regulator proteins with neighbour proteins from protein-protein interaction (PPI) network.

2.2.3.1 Sources for PPI network: BioGRID and IntAct

For mapping my collected list of genes to a common UniProt identifier, UniProt.org Mapper tool was used (http://www.uniprot.org/uploadlists/ (UniProt Consortium, 2014) (conversations made in May 2014). In order to map general protein names the Ensembl Gene database, (Cunningham et al., 2014) version 77 was used *via* the BioMart tool (http://www.ensembl.org/biomart/martview) (Kinsella et al., 2011). Each of the entries has only one unique UniProt identifier, while in some cases there are duplicates because of the concurrent appearance of the reviewed (SwissProt) and un-reviewed (TrEMBL) identifiers.

For PPI connections two PPI databases were used. First is the BioGRID database (http://thebiogrid.org/) (Stark et al., 2006), version 3.2.115 and second is the IntAct database (http://www.ebi.ac.uk/intact/) (Orchard et al., 2014) version 4.1.4. Both of the databases contain only undirected protein-protein connections. Proteins from the BioGRID database proteins were mapped to UniProt IDs, while IntAct database primary keys already provide UniProt IDs. From BioGRID, only physical interactions were obtained, while from IntAct, all of the interactions were used. Finally, a database was established containing only those PPI connections which appear in both databases.

2.2.3.2 Visualization of the network

First neighbour proteins (FNPs) of cell size regulator proteins were added to the network, but only those proteins, which have at least two PPI connections (two degrees[2] connections) to known cell size control proteins (SCPs) (Figure 2.2-1). Five different networks were created based on the five-different species that I selected earlier (*A. thaliana, D. melanogaster, H. sapiens, S. cerevisiae* and *S. pombe*).

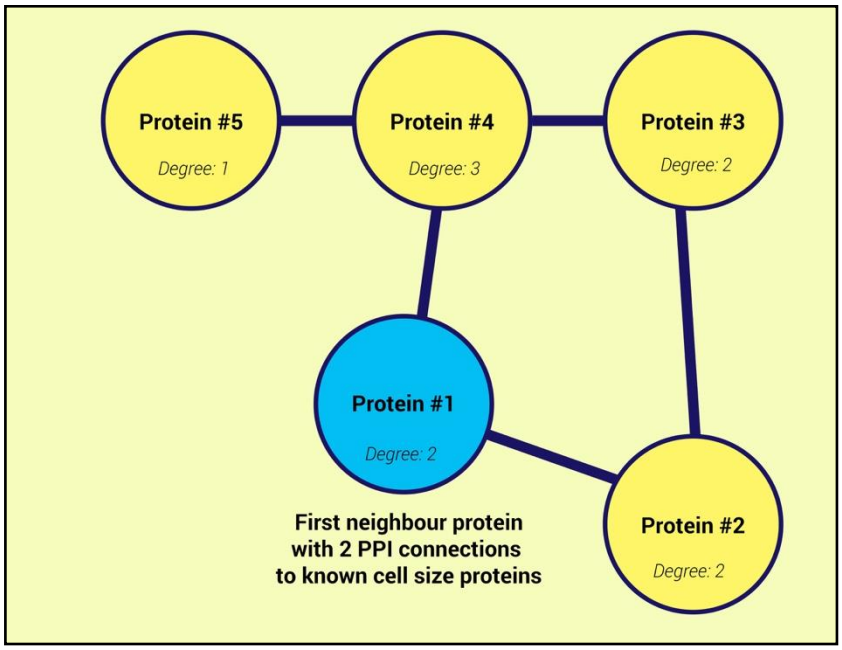


Figure 2.2-1: First neighbour selection criterion.

Protein 2, 3, 4 and 5 are proteins (nodes) encoded by known cell size control genes, while Protein 1 is a first neighbour protein if it is not a cell size regulator, but it has two protein-protein interaction (PPI) connections (edges) to cell size control proteins. **Degree** term describes how many other proteins are connected via PPI connection to a given protein.

The Cytoscape network analyser program was used to visualise PPI networks and calculate the network statistics, (http://www.cytoscape.org/ (Shannon et al., 2003), used version 3.1.1). Data was subsequently exported to CSV files and ordered by Microsoft Excel and analysed by self-made PHP scripts (scripts are available upon request).

[2] **Degree:** The degree of a node in a network means how many edges the node has. Here it means how many first neighbours a particular protein has in the network.

2.2.4 Ortholog and pathway annotation databases

2.2.4.1 Ortholog databases

Ortholog and pathway annotation databases were used to curate ortholog relations. After an investigation of all resources (discussed above), three databases were selected:

(1) eggNOG (Jensen et al., 2008; Powell et al., 2014) version 4.0 was used,
(2) inParanoid (O'Brien et al., 2005; Ostlund et al., 2010) version 8.0 was used,
(3) and a manually curated orthologous list by PomBase curators between *S. pombe* – *S. cerevisiae* and S. *pombe* – *H. sapiens* (retrieved from ftp://ftp.pombase.org/pombe/orthologs/, in May 2014) (Wood et al., 2012)

For a collection of orthologs COG (Clusters of Orthologous Groups for Eukaryotes) groups were used as eggNOG database provided it. I have also noted if an eggNOG suggested orthology relation is backed up by data from InParanoid database.

2.2.4.2 Pathway databases

Pathway enrichment from **KEGG** (Ogata et al., 1999) and **REACTOME** (Croft et al., 2014) was added to each of the proteins in the developed database to be able to search for conserved pathways involved in cell size regulation.

2.2.5 A database for conserved size control proteins: OrthologFinderTool

Separate database files were created for a comprehensive database, using Microsoft Excel, PHP and R scripts (appropriate script files are available upon request). These back-end database files were put together to form a bioinformatics website. The website is based on Apache engine using PHP scripts in the back-end, while the front-end uses a JQuery JavaScript library and the table viewer is based on the DataTables JQuery table plugin (http://www.datatables.net/).

The website can be easily handled *via* a form-based query page. The source code of the whole website can be retrieved from https://github.com/ZoliQua/Ortholog-Finder-Tool/ GitHub address, where a readme document describes how to use the tool and the code.

2.2.6 A database for conserved biological functions: GO Orthology Tool

2.2.6.1 Gene Ontology source

To generalise the orthology search tool I have replaces the data source of cell size regulators with data from the Gene Ontology Database (Ashburner et al., 2000). I used the generic **GO_SLIM** dataset, developed by the GO Consortium (retrieved from http://www.geneontology.org/page/go-slim-and-subset-guide, on 10 November 2015) to reduce the number of GO terms that can describe all biological processes. This list contains 71 generic biological processes (for the whole list of terms see Supplementary Table 6.2-1). I used the proteins belonging to these 71 terms in seven species to find the conserved genes involved in each biological process to create a database for GO Ontology Tool. I created a comprehensive MySQL powered database, that includes a curated list of size control proteins (merged data of collected data explained earlier in Section 2.2.2), together with their eggNOG database and GO_SLIM database annotations. Data from seven organisms including data from different strains (*A. thaliana*, *C. elegans*, *D. melanogaster*, *D. rerio*, *H. sapiens*, *S. cerevisiae* and *S. pombe*) were used with the following taxon ids (Table 2.2-2).

Table 2.2-2: Taxon ids used from Gene Ontology database

Taxon ID	Species Name
3702	*A. thaliana*
6239	*C. elegans*
7227	*D. melanogaster*
7955	*D. rerio*
9606	*H. sapiens*
4932	*S. cerevisiae*
545124	*S. cerevisiae*
559292	*S. cerevisiae*
643680	*S. cerevisiae*
721032	*S. cerevisiae*
764097	*S. cerevisiae*
4896	*S. pombe*

2.2.6.2 Creating Venn Diagrams

I drew Venn Diagrams and Edwards-Venn Diagrams based on samples with Adobe Illustrator software to visualise data that I put it in the cross-species database. Various Venn-Diagrams were produced, ranging from 2 sets to 7 sets (Figure 2.2-2) in SVG (Scalable Vector Graphics) format. I handled these diagrams with scripts (available upon request) to replace letters with relevant numbers.

2.2.6.3 GO Orthology Tool

The database was employed by PHP code on a website back-end, while the front-end was empowered by a JQuery JavaScript library and the table viewer is based on the DataTables JQuery table plugin (http://www.datatables.net/). The Tool handled the manually created Venn Diagrams to show multiple relations among species.

The website can be easily handled *via* a form-based query page. The source code of the whole website can be retrieved from https://github.com/ZoliQua/GO-Orthology-Tool GitHub address, in parallel with a readme document.

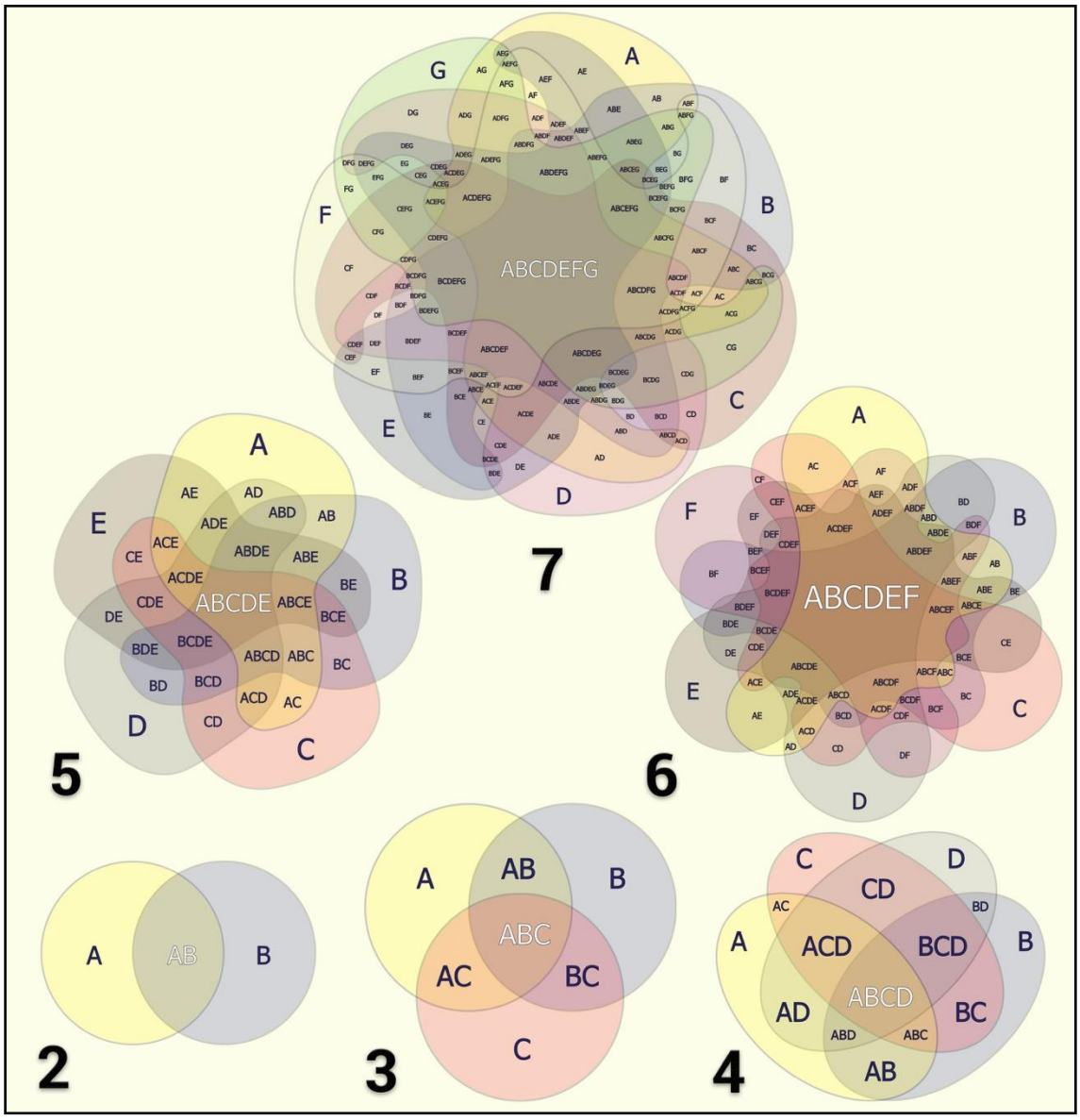


Figure 2.2-2: Venn Diagrams from 2 to 7 sets.

These Venn Diagrams (VDs) were made by using Adobe Illustrator. Pictures were exported in SVG format to easily handle with algorithms. 2 set VD has 3 relations, 3 set VD has 7 relations, 4 set VD has 15 relations, 5 set VD has 31 relations, 6 set VD has 63 relations, 7 set VD has 127 relations.

2.3 Results

2.3.1 The quest for cell size regulators – building a database

I hypothesized that it is possible to uncover the conserved regulatory pathways and find novel regulators of cell size through a comprehensive overview of the known cell size mutants of 5 evolutionally distant species. To reach a broad and comprehensive overview, I created a web-based database of cell size regulators using sources as described in Section 2.2.2 . Using this database, the tool can predict novel affectors of the cell size and identify the universal core regulators of cell size.

2.3.1.1 Creating a database of orthologous proteins

2.3.1.1.1 S. cerevisiae

To obtain budding yeast genes that are involved in the regulation of cell size, data on 451 mutants were obtained from Jorgensen *et al.* (Jorgensen et al., 2002). A name and *Saccharomyces cerevisiae Genome Database* (SGD) identifier correction were performed using the Yeast Genome Database, which reduced the list to 445 genes. From the Moretto *et al.* study (Moretto et al., 2013), 44 new genes were added as mutations caused significant size differences. Among these 44 new cell size regulators 30 were ribosomal proteins.

Overall in budding yeast 489 genes were collected with unique SGD identifiers. These genes were mapped to 490 unique UniProt protein identifiers (UID). From their PPI network, 1,072 first neighbour proteins (FNP) were added. In total, a PPI network with 1,562 size control proteins and 5,514 connections was created (see Figure 2.3-1).

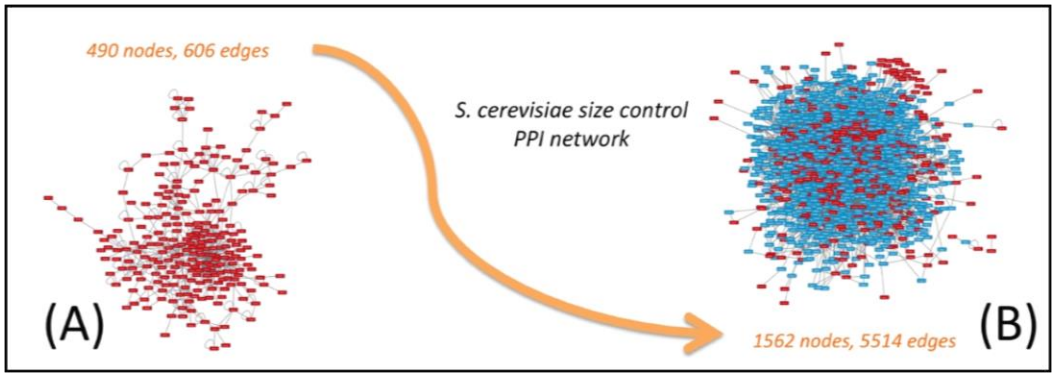


Figure 2.3-1: Two PPI networks of S. cerevisiae

(A) The PPI network of Size Control Proteins (SCPs) in *S. cerevisiae* **(B)** The PPI network of SCPs and First Neighbour Proteins (FNPs) in *S. cerevisiae*. Red nodes are SCPs, while blue nodes are FNPs. Single nodes are not displayed. Network visualisation was created by Cytoscape version 3.1.1.

2.3.1.1.2 S. pombe

To obtain a list of cell size regulators in fission yeast, 871 genes were identified from the Hayles *et al*. study (Hayles et al., 2013), while 90 new genes came from the list of proteins affecting cell size distribution from (Graml et al., 2014). These were merged into 909 genes with unique PomBase Systematic Identifiers (ORF) and then mapped to 932 unique UIDs. From their PPI network, 74 new FNPs were added. In total the network of *Schizosaccharomyces pombe* contained 1,006 proteins with 505 connections. The lower number of connections is a result of the lack of systematic protein interaction studies on fission yeast.

2.3.1.1.3 H. sapiens

To obtain the list of human genes involved in controlling cell size, a query was performed using the MitoCheck primary screen data (Hutchins et al., 2010) for the '*large*' search term. This query identified 316 genes with unique Ensembl gene (ENSG) identifiers. These genes were mapped to 1,051 unique UIDs. From their PPI network, 157 new FNPs were added. In total the network of human cell size related proteins contained 1,208 proteins with 431 connections. Unfortunately, there was no data on small cells, so in this case I focused only on mutants with larger cell size.

2.3.1.1.4 D. melanogaster

To obtain the list of Drosophila genes which control cell size, 385 genes were retrieved from the Björklund *et al*. study (Björklund et al., 2006) with unique FlyBase genome (FBgn) identifiers. These genes were mapped to UIDs.

Also, the FlyBase database (St Pierre et al., 2014) was queried for the following terms:

- '*size defective*',
- '*increased cell size*',
- '*decreased cell size*' and
- '*body size defective*'.

These queries contained 1,574 unique UIDs in the results. Then results were merged with the Bjorklund study into 2,193 unique UIDs. From their PPI network, 860 new FNPs were added.

In total the network of Drosophila cell size related proteins contained 3,053 proteins with 3,376 connections.

2.3.1.1.5 A. thaliana

There were no published genome-wide studies on gene mutants affecting cell size for *A. thaliana*. Therefore the Gene Ontology (GO) database (Ashburner et al., 2000) was used to query loss-of-function defective mutants. The GO database was queried for the GO term GO:0008361 '*regulation of cell size*'.

As a result of the query 57 genes were obtained. From these 55 have unique *The Arabidopsis Information Resource* (TAIR) identifiers (Huala et al., 2001). TAIR identifiers were mapped to 68 unique UIDs. From their PPI network, only one FNP was added. In total the network of *Arabidopsis* cell size related proteins contained 69 proteins with 4 connections.

2.3.1.2 PPI network, orthologous and pathway annotation databases for cell regulators database

In each of the organisms, two PPI networks were created: one restricted to known cell regulators and one expanded network with added FNPs. In total 10 PPI networks were created, then all data were exported to separate spreadsheets and merged into one database file per organism. To obtain orthologous proteins of these species, three different orthologous sources were integrated, and the biological pathway annotation was applied.

2.3.1.3 Comprehensive Database - Ortholog Finder Tool

I created a website for the database with a user-friendly front-end (Figure 2.3-2). The result of a query appears straightaway after every single query from the database file sources (Figure 2.3-3). The tool can be reached at the http://www.orthologfindertool.com/ web address. (Note: in Figure 2.3-3, data for Table 2.3-1, **row B**).

Each results page can be downloaded as a CSV (comma separated value) file for further analysis.

Figure 2.3-2: Ortholog Finder Tool: Front-end of the query page

Users can select from one option for the Dataset and Function as shown, while there are five species available.

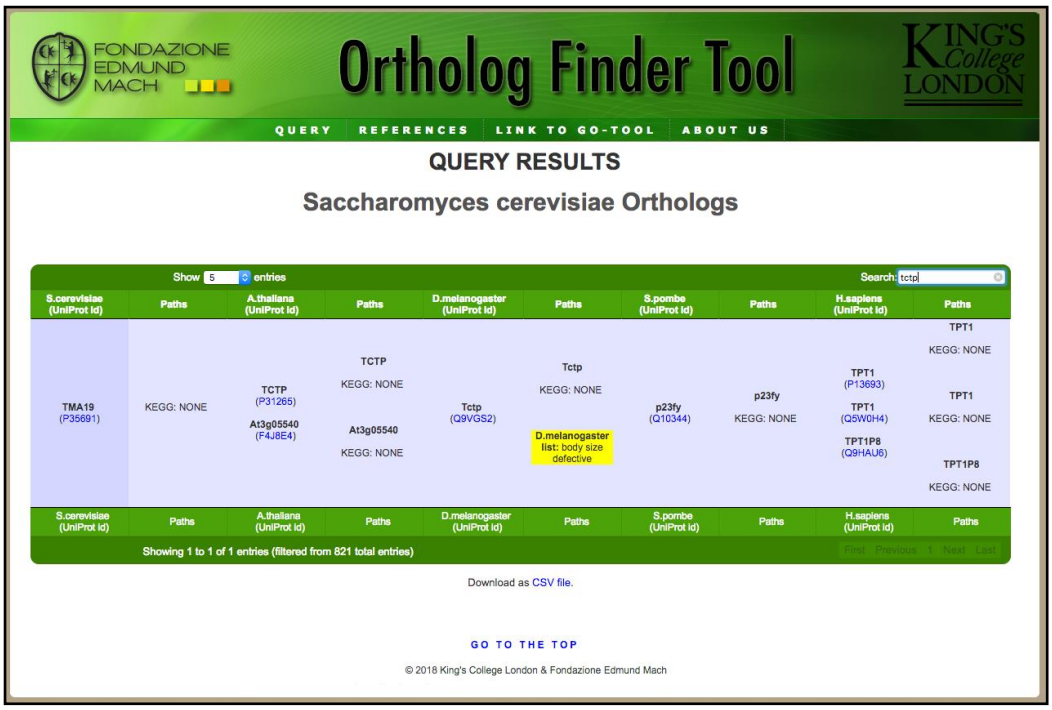


Figure 2.3-3: Ortholog Finder Tool: Front-end of the results page

Here I show the query/result page of a query taken on the *S. cerevisiae* database and then filtered to "*tctp*" term. TCTP in *Drosophila* is an ortholog of the TMA19 budding yeast protein (Rinnerthaler et al., 2006) and TPT1/TCTP in human. The appearance of proteins in other SCP lists is always marked by a yellow background text. I also provide pathway enrichment from KEGG for each protein. However, in this given example none of the proteins had one.

2.3.2 Selecting genes for experiments

As one of my main aims was to find novel cell size regulators in *Arabidopsis* and human, I began by analysing the database with the idea to find proteins that are conserved in these organisms and also in the investigated two yeasts and in *Drosophila*. I was specifically looking for cases when the yeast and *Drosophila* orthologs were already associated with cell size regulation, but this was not shown in human and Arabidopsis. The resulting lists were exported from the database then systematically analysed.

1. Initially, I reduced query results to those proteins, which have at least one orthologous protein in one of the other organisms (Table 2.3-1, **row A**) (e.g. a budding yeast protein has a fission yeast ortholog).
2. Then to those proteins which have an ortholog in all five organisms (Table 2.3-1, **row B**).
3. Subsequently, I reduced these lists to only include those proteins which have a pathway annotation (Table 2.3-1, **row C**)

Table 2.3-1: Merging and reducing the number of candidate proteins

	A. thaliana[3]	*D. melanogaster*	*H. sapiens*	*S. cerevisiae*	*S. pombe*
Size-control proteins (SCPs)	68	2193	1051	490	932
First neighbour proteins (FNPs) of SCPs	1	860	157	1072	74
SCPs + FNPs	**69**	**3053**	**1208**	**1562**	**1006**
(A) Proteins that have at least one orthologous protein	24	2207	356	1398	923
(B) Proteins that have at least one orthologous protein in all other species	**13**	**567**	**139**	**821**	**512**
(C) Proteins that have at least one orthologous protein in all	5	328	76	479	294

[3] *Arabidopsis* numbers are significantly lower because of the significantly lower amount of source data.

other species and has pathway annotation					

4. Then, I merged the five lists of databases into one grand database, which I started to reduce further. I began to select those orthologous groups, where there was more than one evidence of cell size regulation. It means that out of *S. cerevisiae*, *S. pombe* and *D. melanogaster* at least two of them are either reported SCP, or one reported and one of them is an FNP of at least two reported SCPs.
5. Later, ribosomal proteins were removed since perturbations of ribosomes could lead to changes in the rate of translation, the cellular growth rate and therefore cell size (Jorgensen & Tyers, 2004; Jorgensen et al., 2002). Still, they are not directly involved in the regulation of cell size and cell volume. These were removed from my analysis since they could have an indirect and common role in cell size regulation and their sheer number could bias subsequent analyses. Probably other groups of proteins could have been removed for the same reasoning, but no other group had such a distinct role in growth regulation.
6. From this list, I removed protein groups that have been shown to control cell size in human and *Arabidopsis*. However, I used some of them as a positive control (such as PFD6). After this point the list contained proteins that were not known to be involved in cell size regulation either in human or *Arabidopsis*.
7. Later, I developed a method to check the number of articles mentioning the candidate proteins and cell size. I named it "*PubMed score*". It means that if someone puts a query into `https://www.ncbi.nlm.nih.gov/pubmed` with the name of the protein and species it will report the same number of academic papers as the "*PubMed score*". Accordingly, I have removed those proteins which had a PubMed score of 6 or more to focus on less investigated candidate proteins.
8. As a result, I selected about 40 proteins that I further investigated by manual checking. Following these checks, I created a list in which only those orthologous groups were considered where only one protein is present in each organism.
9. The final list contained 17 proteins (Table 2.3-2) and from this list I chose 5 proteins. I have chosen these candidates because they are known to regulate cell

size in at least one organism and there is no information on their roles in controlling cell size in human and *Arabidopsis*.

Table 2.3-2: Table of one-to-one ortholog proteins investigated

The column marked "X" means the orthologous group was selected for the experimental test.

Selected	A. thaliana	D. melanogaster	H. sapiens	S. cerevisiae	S. pombe
	ARP6	Arp6	ACTR6	ARP6	arp6
X	At1g12470	dor	VPS18	PEP3	pep3
	At1g31780	CG1968	COG6	COG6	cog6
	At1g42440	CG7338	TSR1	TSR1	tsr1
	At1g69070	l(3)07882	NOP14	NOP14	nop14
X	At5g26110	Prpk	TP53RK	BUD32	bud32
X	At5g49510	mgr	VBP1	PAC10	pac10
X	At5g62590	Trn-SR	TNPO3	MTR10	mtr10
	F5M6.12	CG13625	BUD13	BUD13	cwf26
	F6I18.100	CG4554	UTP20	UTP20	utp20
	INO80	Ino80	INO80	INO80	ino80
	MCM23.1	CG9004	NOM1	SGD1	sgd1
	MED31	MED31	MED31	SOH1	med31
	MED6	MED6	MED6	MED6	med6
	MFB13.10	Aatf	AATF	BFR2	bfr2
	SWC4	DMAP1	DMAP1	SWC4	swc4
X	TCTP[4]	TCTP	TCTP	TMA55	tma19

SCP FNP

[4] This information has been revealed by a later literature review (Berkowitz et al., 2008).

I selected TCTP/TPT1, TP53RK, VBP1, VPS18 and TNPO3 as candidates for wet-laboratory experiments. The other candidates were dropped for various reasons, including that their functions were already known, the gene is crucial for cell maintenance (the knock-out or knock-down of the protein was lethal), and/or the deletion or silencing of the gene was not reported. Thus, I have selected five candidate proteins, which I investigated by gene deletion in *Arabidopsis* and by siRNA experiments using human T cells isolated from peripheral blood.

2.3.2.1 Orthologs of TCTP/TPT1

Translationally Controlled Tumour Protein (**TCTP**) is a highly conserved, multifunctional protein that is involved in several fundamental biological processes either in human or other species (Bommer, 2017). It has only one copy in all five species that I investigated, which suggest it could have a dosage related function. One copy means in each of the species it fulfils a specific role, and this could explain why this gene was not duplicated in any of these species. Moreover, it has only one copy in *C. elegans* and *D. rerio* as well. The homologous gene in *Drosophila* has been described as a cell size and body size defective mutant (Björklund et al., 2006), while in budding yeast it is an FNP protein (see Table 2.3-3) (PHO85 (YPL031C), RPN11 (YFR004W) and UBP3 (YER151C)). Deletion of *TCTP* in *Arabidopsis* causes reduced cell size (Berkowitz et al., 2008). However, I was not able to reproduce this phenotype in my *Arabidopsis* experiments (see Section 4.3.4).

Table 2.3-3: Orthologs of the TCTP protein and current knowledge of its involvement in cell size regulation.

Species	*Protein name*	*Involvement in size regulation*
A. thaliana	*TCTP [AT3G16640]:*	*P31265:* ***Reduces cell size*** *(Berkowitz et al., 2008)*
D. melanogaster	*TCTP*	*Q9VGS2:* ***SCP: cell size defective, decreased cell size, size defective, body size defective*** *(Björklund et al, 2006)*
H. sapiens	*TPT1*	*P13693*
S. cerevisiae	TMA55 (YKL056C)	*P35691:* ***FNP***
S. pombe	*tma19 [SPAC1F12.02c]*	*Q10344*

2.3.2.2 Orthologs of TP53RK/Bud32

TP53RK, also known as Bud32 and PRPK, is a transcriptional regulator and tumour suppressor gene that regulates the phosphorylation of TP53 on Serine 15 (Miyoshi et al., 2003). TP53RK protein has one homologue in all seven organisms (incl. *C. elegans* and *D. rerio*). The gene encodes a cell size control protein in budding yeast (Jorgensen et al., 2002) and *Drosophila* (Björklund et al., 2006). See Table 2.3-4.

Table 2.3-4: Orthologs of the TP53RK protein and current knowledge of its involvement in cell size regulation.

Species	Protein name	Involvement in size regulation
A. thaliana	*At5g26110*	*Q94K14*
D. melanogaster	*CG10673 [FBgn0035590]*	*Q9VRJ6:* ***SCP: cell size defective, decreased cell size, size defective, body size defective*** *(Björklund et al, 2006)*
H. sapiens	*TP53RK*	*Q96S44*
S. cerevisiae	BUD32 (YGR262C)	*P53323:* ***SCP****:* ***haploid-larger*** *(Jorgensen et al., 2002)*
S. pombe	*bud32 [SPAP27G11.07c]*	*Q7P7N1*

2.3.2.3 Orthologs of VPS18/PEP3

VPS18 (Vacuolar protein sorting-associated protein 18 homolog) protein in human plays an important role in the vesicle-mediated protein trafficking to the lysosome (Wurmser et al., 2000) and the protein is also required for vacuolar biogenesis in yeast (Preston et al., 1991). The protein is a part of HOPS (homotypic fusion and vacuole protein sorting) and CORVET (class C core vacuole/endosome tethering) complexes, while it controls the fusion of endomembrane (Hunter et al., 2017). VPS18 was found to control cell size in fission and budding yeast (Hayles et al., 2013; Jorgensen et al., 2002), while it is an FNP in the *Drosophila* network. In all of the seven species analysed only one copy of the gene is present.

2.3.2.4 Orthologs of TNPO3

TNPO3 (Transportin 3) protein is a receptor located in the nuclear membrane where it is responsible for transporting serine/arginine-rich proteins such as the splicing factors RBM4, SFRS1 and SFRS2 (Allemand et al., 2002; Kataoka et al., 1999). TNPO3 is involved in HIV-1 infections, while it interacts with capsid proteins (Krishnan et al., 2010). The TNPO3 protein ortholog in fission yeast has been identified as a cell size regulator (Hayles et al., 2013), while budding yeast and fly orthologs have been described as FNP proteins.

2.3.2.5 Orthologs of VBP1/PFD3

VBP1 (von Hippel-Lindau-binding protein) protein in humans is a component of the prefoldin complex and It Is also known as Prefoldin 3.The Prefoldin complex promotes cellular protein folding into the proper conformation and maintains homeostasis (Siegert et al., 2000). Apart from its involvement in the protein folding, VBP1 is a member of the ubiquitin protease system and stimulates protein degradation (Mousnier et al., 2007; Tsuchiya et al., 1996). VBP1 protein orthologs in budding yeast (Pac10) and fission yeast have been identified as cell size regulators (Hayles et al., 2013; Jorgensen et al., 2002), in line with the *Drosophila* ortholog that has been described as an FNP protein. One copy of the VBP1/PFD3 gene is present in all of the species investigated and the same is true for genes encoding the other components of the prefoldin complex.

2.3.2.6 Selection of genes of the prefoldin complex orthologs

After carrying out the first round of gene deletion experiments on *Arabidopsis* I have identified *VBP1/PFD3* as a good candidate for further research (see Section 4.3.3.1). Since this protein is a member of the prefoldin complex, I decided to investigate other components of this complex (see Table 2.3-5). These proteins form a heterohexameric molecular complex that is highly conserved in eukaryotes, including humans (Cao, 2016). In the complex, there are two alpha (PFD3 and PFD5) and four beta subunits (PFD1, PFD2, PFD4 and PFD6) (Siegert et al., 2000). The prefoldins, in conjunction with molecular chaperonins are mainly involved in the correct folding of cellular proteins (Vainberg et al., 1998).

Table 2.3-5: Members of the prefoldin complex.

	PFD1/PFDN1	***PFD2/PFDN2***	***PFD4/PFDN4***	***PFD5/PFDN5***	***PFD6/PFDN6***
A. thaliana	***At2g07340***	***At3g22480***	***At1g08780***	***At5g23290***	***At1g29990***: *decreased size, microtubular defect*
D. melanogaster	***Pfdn1*** *[FBgn0031776]*	***Pfdn2*** *[FBgn0010741]*	***Pfdn4*** *[FBgn0035603]*	***Pfdn5*** *[FBgn0038976]*	***Pfdn6*** *[FBgn0036918]*
H sapiens	***PFDN1*** *[O60925]*	***PFDN2*** *[Q9UHV9]*	***PFDN4*** *[Q9NQP4]*	***PFDN5*** *[Q99471]*	***PFDN6*** *[O15212]*
S. cerevisiae	***GIM6*** *FNP, increased cell size*	***GIM4*** *FNP, increased cell size*	***GIM3*** *FNP, increased cell size*	***GIM5*** *increased cell size*	***YKE2*** *increased cell size*
S. pombe	***gim6*** *[SPBC1D7.01]*	***gim4*** *[SPAC227.10]*	***gim3*** *[SPAC227.05]*	***bob1/gim5*** *[SPBC215.02]*	***gim1*** *[SPAC3A11.13]*

2.3.2.7 Orthologs of CDC7

In addition to the prefoldin complex, I selected another orthologous group for wet-laboratory experiments. CDC7 in humans is a protein kinase that is important for the G_1/S transition in the cell cycle (Montagnoli et al., 2002). The protein binds chromatin and interacts with MCM2 and MCM4 proteins to regulate the activity of the MCM hexamer DNA replication complex (Masai et al., 2006; Tsuji et al., 2006). CDC7 was characterised as a cell size regulator protein in the fly (Björklund et al., 2006) and fission yeast (Hayles et al., 2013), while in budding yeast it has been described as an FNP.

Table 2.3-6: Orthologs of CDC7 and their involvement in size regulation.

Species	Protein name	Involvement in size regulation
A. thaliana	***At4g16970***	*Q0WPK0*
D. melanogaster	***Cdc7*** *[FBgn0028360.]*	*Q9W3Y1:* ***SCP: cell size defective, decreased cell size, size defective, body size defective*** *Q8T0Q5 and Q9VJ90*
H. sapiens	***CDC7***	*O00311*
S. cerevisiae	**CDC7** (YDL017W)	*P06243:* **FNP**
S. pombe	***hsk1*** *[SPBC776.12c]* ***spo4*** *[SPBC21C3.18]*	*P50582:* ***SCP: long high penetrance*** *(Hayles et al.)* *Q9UQY9*

These orthologs have an evident involvement in cell cycle regulation and as Table 2.3-6 shows they also affect cell size in some organisms. Thus, I wanted to test the effects of perturbations of CDC7 on cell size in human T-cells and *Arabidopsis*.

2.3.3 Gene Ontology Orthology Tool (GOOT)

Based on the orthology of proteins that are associated with a common biological process it Is possible to search for the conserved set of genes that could be important for a given biological process in selected species. Thus, I created a tool that can rank eggNOG orthologous groups by the conservation of GO annotated biological processes in which their members are involved. The same tool can also be used to identify gaps in GO annotations. Since, if I find that an orthologous group contains six proteins in six investigated species that all belong to the same GO annotation, then it is likely that a seventh group member from a seventh species is also involved in the same biological process.

To fulfil these aims I built an extension of the tool called the **Gene Ontology Orthology Tool (GOOT)** to analyse the current status of gene ontology functional terms based on orthology. This tool shows the conserved function of proteins for a given GO term and gives suggestions for novel functional annotations based on data annotated for other species.

2.3.3.1 Building up the database - sources

To build the tool, I used external datasets from the following databases:

(1) eggNOG database (version 4.0) was used for orthology

(2) GO_SLIM database was used for GO terms, which contains the list of 71 generic biological processes

   a. The list of GO terms included are given in Supplementary Table 6.2-1

   b. The dataset was last updated on 10th August 2017.

I created a website for the tool that is up and running on the web. The address of the website is http://go.orthologfindertool.com/.

2.3.3.2 Using the GOOT website

The interface of the website welcomes the user; then users can enter a query (Figure 2.3-4). There are four options on the query page:

- Selection from the 71 GO Slim terms.
- Selection from seven organisms (*A. thaliana*, *C. elegans*, *D. melanogaster*, *D. rerio*, *H. sapiens*, *S. cerevisiae* and *S. pombe*) - multiple choices are possible.
- Selecting from two types of Venn Diagrams (normal and Edwards-Venn diagram).
- Determine a hit species threshold level between one and seven, meaning how many of the seven species need to have a protein from a given orthologous group associated with the selected GO term.

The **hit species threshold** level means that an orthologous group needs to have an annotated protein in the GO in the threshold number of species. If the threshold number is three and we select "*GO:0051726 - regulation of cell cycle*" GO term, there are 103 orthologous groups that have a protein, in at least three species annotated to this GO term. 45 of the orthologous groups are present in all seven organisms. Just to compare, if the threshold number is one as shown in Figure 2.3-5, there are 244 common orthologous groups of proteins that have at least one annotation to this GO term and are present in all seven species. The value of hit species threshold is always 1 for the Venn Diagrams shown in this chapter.

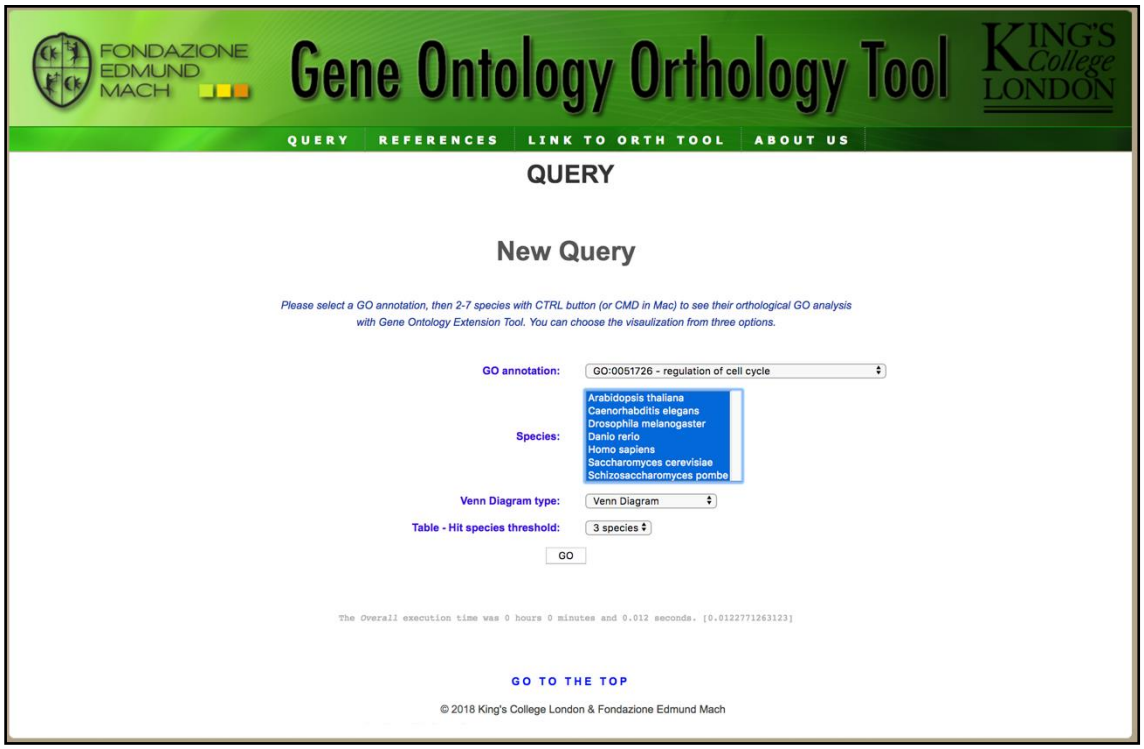


Figure 2.3-4: Query page of the Gene Ontology Orthology Tool

Users can select from 71 GO terms, seven species, two types of Venn diagrams and seven threshold levels.

2.3.3.3 Results

2.3.3.3.1 Venn Diagrams

Once the user sends a query, the results page will load a Venn diagram of the selected GO term at the top of the page (Figure 2.3-5). It describes all the possible orthological overlaps among the selected number of species, then represents the number of common orthologous groups in each possible relation. Each number in the diagram means one single orthologous group that contain several unique proteins in species, from one protein to hundreds of proteins. The numbers show the orthologous groups that have at least one protein out of the seven species annotated in the selected GO term (the hit species threshold number is always 1 for Venn Diagrams).

In Figure 2.3-5, I show three different Venn diagrams for the "*GO:0051726 - regulation of cell cycle*" that can be created with the tool, depending on which species are selected for analysis.

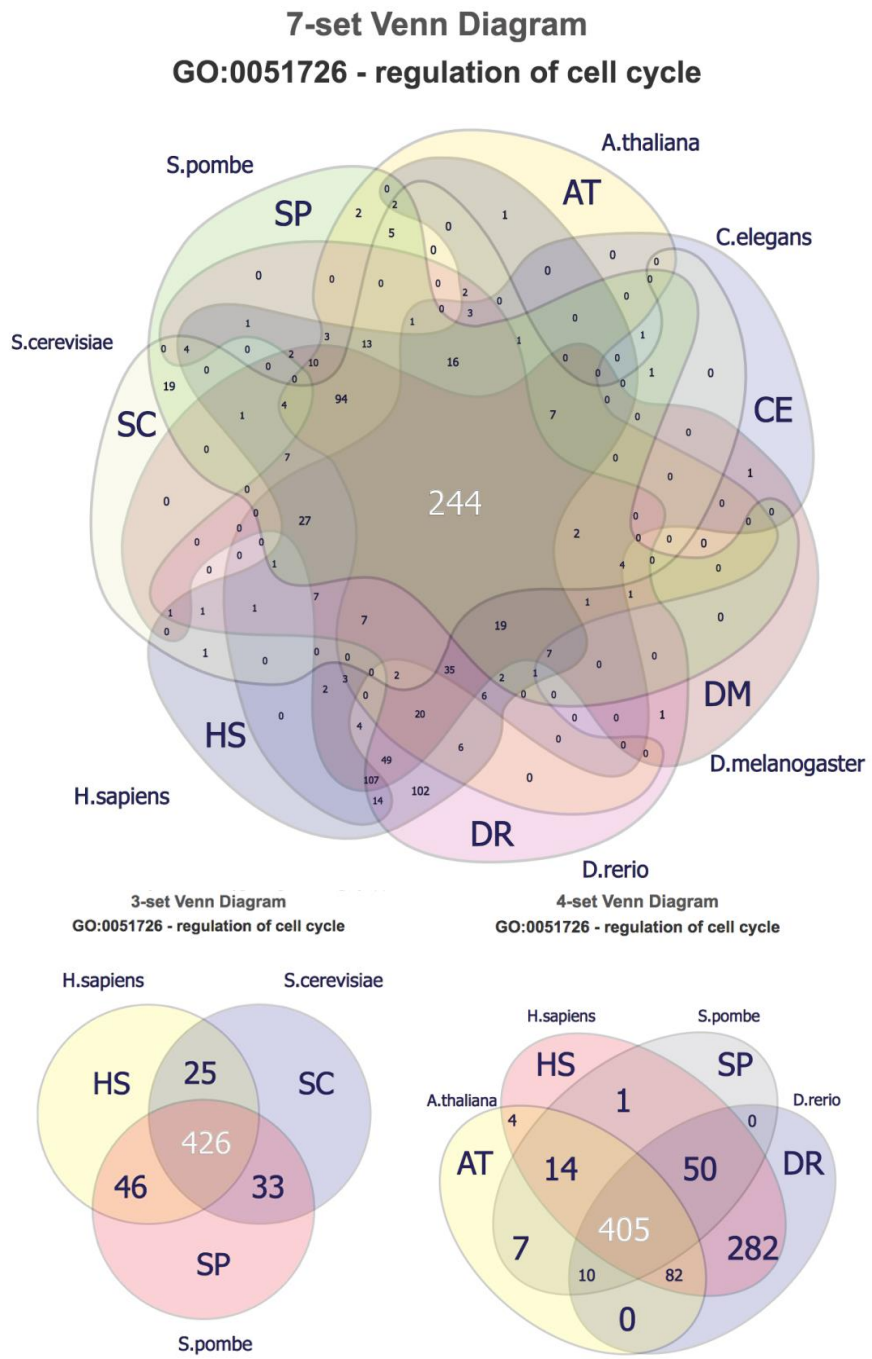


Figure 2.3-5: Regulation of Cell Cycle – Venn Diagrams

"*GO:0051726 - regulation of cell cycle*" GO term and three different queries: among seven-, among four- and among three-species. Each number denotes an orthologous group from eggNOG that contains proteins from the given species for each overlap. Only overlapping groups are displayed, species-specific orthologous groups are not. There are 426 orthologous groups (bottom left) where there is at least one gene from budding yeast, fission yeast and humans annotated to this GO term.

2.3.3.3.2 Current Annotations based on orthology

On the result page shown by the tool, the second section is the current knowledge functions of Gene Ontology annotations based on orthology. Here the tool shows a table of the annotated functional orthologous groups above the selected hit species threshold level.

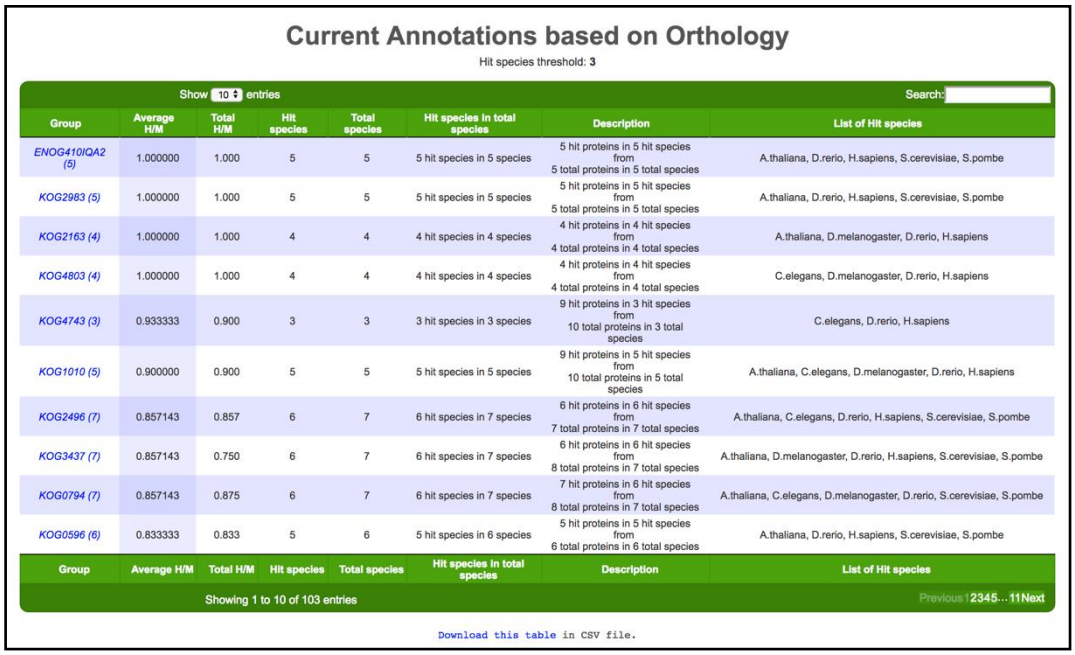
The table contains the following columns (see Figure 2.3-6):

- **Group:** eggNOG group name that links to the related eggNOG group on the website of the eggNOG database
- **Average H/M**: Individual Hit/Member (H/M) ratio of the species calculates the number of given GO annotated proteins (**hit**) divided by all proteins (**member**) annotated to that eggNOG orthologous group in the given species. A calculated average number makes an average of individual hit/member ratios of individual species, then divides it by the number of species present in that orthology group.

  I needed to consider this as some genes have many copies in a single organism, while present as a single copy in others. I did not want to penalise cases when the biological function might be conserved for one of the many copies in one species as well as in the single copies in other species.
- **Total H/M**: Total Hit/Member ration is a summarized number of annotated proteins to that term (**hit**) divided by the number of all proteins (**member**) exist in the orthologous group. This measure can be misleading if there are multiple copies of a gene in one or two species while a few copies in others.
- **Hit species:** The number of species that have at least one protein annotated to the given GO term.
- **Total species:** The number of all species that have at least one protein in the orthologous group.
- **Hit species in total species**: Describes the number of species that have a protein annotated (hit species) and the number of all species (total species) in the orthologous group.

- **Description**: Describes the number of annotated proteins in the number of hit species and describes the number of all proteins in all species that belongs to this eggNOG orthologous group.
- **List of hit species**: List of species that have a protein annotated to the given GO term in the orthologous group.

The Average H/M ratio is able to discriminate among groups to show the most relevant GO functional orthologous groups. As Figure 2.3-6 shows "*GO:0051726 - regulation of cell cycle*" GO term orthologous groups ordered by Average H/M in descending order. We can discriminate a conserved core function of proteins where the total H/M value is 1 and the number of hit species equals the total number of species in the databases used by the tool. There are four orthologous groups (ENOG410IQA2, KOG2983, KOG2163 and KOG4803), with four proteins that met this criterion. Furthermore, if we set total H/M value to 0.9, while the hit species still equals to the total species, we can discriminate an additional two orthologous groups, KOG4743 and KOG1010, increasing the number of identified conserved orthologous groups to six.
In general, this tool can be used to rank eggNOG groups based on their association in multiple organisms to a given GO term. Thus, it can be used to identify the most conserved genes associated with a given ontology association.



Current Annotations based on Orthology

Hit species threshold: **3**

Show 10 entries Search:

Group	Average H/M	Total H/M	Hit species	Total species	Hit species in total species	Description	List of Hit species
ENOG410IQA2 (5)	1.000000	1.000	5	5	5 hit species in 5 species	5 hit proteins in 5 hit species from 5 total proteins in 5 total species	A.thaliana, D.rerio, H.sapiens, S.cerevisiae, S.pombe
KOG2983 (5)	1.000000	1.000	5	5	5 hit species in 5 species	5 hit proteins in 5 hit species from 5 total proteins in 5 total species	A.thaliana, D.rerio, H.sapiens, S.cerevisiae, S.pombe
KOG2163 (4)	1.000000	1.000	4	4	4 hit species in 4 species	4 hit proteins in 4 hit species from 4 total proteins in 4 total species	A.thaliana, D.melanogaster, D.rerio, H.sapiens
KOG4803 (4)	1.000000	1.000	4	4	4 hit species in 4 species	4 hit proteins in 4 hit species from 4 total proteins in 4 total species	C.elegans, D.melanogaster, D.rerio, H.sapiens
KOG4743 (3)	0.933333	0.900	3	3	3 hit species in 3 species	9 hit proteins in 3 hit species from 10 total proteins in 3 total species	C.elegans, D.rerio, H.sapiens
KOG1010 (5)	0.900000	0.900	5	5	5 hit species in 5 species	9 hit proteins in 5 hit species from 10 total proteins in 5 total species	A.thaliana, C.elegans, D.melanogaster, D.rerio, H.sapiens
KOG2496 (7)	0.857143	0.857	6	7	6 hit species in 7 species	6 hit proteins in 6 hit species from 7 total proteins in 7 total species	A.thaliana, C.elegans, D.rerio, H.sapiens, S.cerevisiae, S.pombe
KOG3437 (7)	0.857143	0.750	6	7	6 hit species in 7 species	6 hit proteins in 6 hit species from 8 total proteins in 7 total species	A.thaliana, D.melanogaster, D.rerio, H.sapiens, S.cerevisiae, S.pombe
KOG0794 (7)	0.857143	0.875	6	7	6 hit species in 7 species	7 hit proteins in 6 hit species from 8 total proteins in 7 total species	A.thaliana, C.elegans, D.melanogaster, D.rerio, S.cerevisiae, S.pombe
KOG0596 (6)	0.833333	0.833	5	6	5 hit species in 6 species	5 hit proteins in 5 hit species from 6 total proteins in 6 total species	A.thaliana, D.rerio, H.sapiens, S.cerevisiae, S.pombe

Showing 1 to 10 of 103 entries Previous 1 2 3 4 5 … 11 Next

Download this table in CSV file.

Figure 2.3-6: Current Annotations based on Orthology

Data for "*GO:0051726 - regulation of cell cycle*" GO term. The Figure shows the first ten rows ordered by Average H/M in descending order. It can be seen that there are four orthologous groups which have all of the proteins annotated with this GO term in all the investigated species. Hit species threshold equals 3.

2.3.3.3.3 Novel GO Annotations

The second table on the results page is the main result of the tool: it suggests 11 novel functional annotations based on orthology. The tool shows orthologous groups which have all but one hit species, Average H/M ratio **0.800** or above, and Total H/M ratio **0.500** or above. These selected orthologous groups are shown in a table (Figure 2.3-7), where all the proteins are shown for all investigated species, in line with the Group name, Average H/M ratio and Total species numbers. In all hit species the "**HIT:**" tag and the novel functional annotation prediction the "***PREDICTION:***" tag is shown. For the "*GO:0051726 - regulation of cell cycle*" GO term, 13 protein predictions from 11 orthologous groups can be made (Figure 2.3-7 and Table 2.3-7)

Novel Annotations based on Orthology

Show All entries Search:

Group	Average H/M	Total species	A. thaliana	C. elegans	D. melanogaster	D. rerio	H. sapiens	S. cerevisiae	S. pombe
KOG2496 (7)	0.857143	7	HIT: Q8W5S1	HIT: Q9N4U0	PREDICTION: O17144	HIT: A2BGI0	HIT: P51946	HIT: P37366	HIT: P36613
KOG3437 (7)	0.857143	7	HIT: Q9ZPW2	PREDICTION: Q2WF61, H2L2B8	HIT: Q9V831	HIT: E7FCJ1	HIT: Q9UM13	HIT: P53068	HIT: O42971
KOG0794 (7)	0.857143	7	HIT: Q9FJK6, Q9FJK7	HIT: Q9TYP2	HIT: P25008	HIT: F1RBC9	PREDICTION: Q7Z4L3	HIT: P47821	HIT: O94503
KOG0596 (6)	0.833333	6	HIT: Q84VX4	-	PREDICTION: A0A0B4KG66	HIT: E9QBU7	HIT: P33981	HIT: P54199	HIT: O94235
KOG4593 (6)	0.833333	6	HIT: Q9LTY1	-	PREDICTION: Q95S25	HIT: D9IWE2	HIT: Q9Y6D9	HIT: P40957	HIT: P87169
KOG1155 (6)	0.833333	6	HIT: Q9STS3	-	PREDICTION: E1JHM5, Q9I7L8	HIT: Q7ZVH3	HIT: Q9UJX2	HIT: P16522	HIT: O94556
KOG3285 (6)	0.833333	6	HIT: Q9LU93	-	PREDICTION: Q9VRQ2	HIT: F1QHN1	HIT: Q13257	HIT: P40958	HIT: O14417
KOG2607 (5)	0.800000	5	PREDICTION: Q9FG23	HIT: Q9U2Y2	HIT: Q95SK3	HIT: F1Q5I0	HIT: Q96JB5	-	-
KOG0835 (5)	0.800000	5	HIT: Q8RWV3	-	PREDICTION: Q9U1K6	HIT: B8JJS9, A2BID3	HIT: A0A024R072, Q9UK58	-	HIT: O94612
KOG2810 (5)	0.800000	5	HIT: Q9MA61	-	PREDICTION: A8JNV5	HIT: F2Z4U5, B0S509	HIT: Q99638, J3KPN7	-	HIT: P26306
KOG3924 (5)	0.800000	5	-	HIT: Q18939, Q23375, Q20856, Q20794, Q23200	PREDICTION: A0A0B4KFP4	HIT: F1Q4W7	HIT: Q8TEK3	HIT: Q04089	-
Group	Average H/M	Total species	A. thaliana	C. elegans	D. melanogaster	D. rerio	H. sapiens	S. cerevisiae	S. pombe

Showing 1 to 11 of 11 entries Previous 1 Next

Download this table in CSV file.

Figure 2.3-7: Novel GO Annotations based on orthology

Data for "*GO:0051726 - regulation of cell cycle*" GO term. The figure shows the 11 rows of orthologous groups where 13 novel functional annotations can be made.

For example, here the tool predicts the **Cyclin H** (O17144) protein from *D. melanogaster* as the "*GO:0051726 - regulation of cell cycle*" GO term, based on KOG 2496 - *Cyclin H - Cell cycle control, cell division, chromosome partitioning* – orthologous group.

Table 2.3-7: Predictions for GO:0051726 - regulation of cell cycle

Predictions based on KOG 2496 - *Cyclin H - Cell cycle control, cell division, chromosome partitioning* orthologous group.

GO:0051726 - regulation of cell cycle				
Group	***Total species***	***Species of Predicted Protein***	***Name of Predicted Protein***	***Uniprot ID of Predicted Protein***
KOG2496	7	*D. melanogaster*	Cyclin H	O17144
KOG3437	7	*C. elegans*	Apc-10, Y48G1C.12	Q2WF61, H2L2B8
KOG0794	7	*H. sapiens*	CCNC	Q7Z4L3
KOG0596	6	*D. melanogaster*	Mps1	A0A0B4KG66
KOG4593	6	*D. melanogaster*	Mad1	Q95S25
KOG1155	6	*D. melanogaster*	CG31687-PA, Cdc23	E1JHM5, Q9I7L8
KOG3285	6	*D. melanogaster*	Mad2	Q9VRQ2
KOG2607	5	*A. thaliana*	CDK5RAP3-like protein	Q9FG23
KOG0835	5	*D. melanogaster*	EG:67A9.2	Q9U1K6
KOG2810	5	*D. melanogaster*	Rad9	A8JNV5
KOG3924	5	*D. melanogaster*	gpp	A0A0B4KFP4

The first protein CycH (O17144) from the **KOG2496** group is informative by its name, it was identified by similarity in eggNOG and the evidence is at the transcriptome level. However, it has not been annotated yet to this GO term, unlike other Cyclin H proteins in other organisms.

2.3.3.4 GO Results for Regulation of cell size

As I created a new tool to find novel functional annotations based on orthology, still my aim remained to find novel cell size regulators. The GO database contained the "*GO:0008361 - regulation of cell size*" GO term. I found that 37 individual orthologous groups have at least one protein annotated in a minimum of one species of the seven-investigated species (Figure 2.3-9, top). Moreover, I found that there are 47 orthologous groups among my initial five species (*A. thaliana*, *D. melanogaster*, *H. sapiens*, *S. cerevisiae* and *S. pombe*) that have at least one protein annotated in at least one species (Figure 2.3-9, bottom right). In addition, I found that there are 70 orthologous groups between *S. cerevisiae* and *S. pombe* that have at least one protein annotated in at least one species (Figure 2.3-9, bottom left).

2.3.3.4.1 Applying threshold levels

I applied hit species threshold level 2, that means a given orthologous group needs to have an identified functional ortholog in at least 2 species. I found that 29 orthologous groups have three proteins annotated in at least three species. When a threshold level of 3 Is applied, I found that 9 orthologous groups have three proteins annotated in at least three species (Figure 2.3-8).

Current Annotations based on Orthology

Hit species threshold: **3**

Show 10 entries Search:

Group	Average H/M	Total H/M	Hit species	Total species	Hit species in total species	Description	List of Hit species
KOG3238 (5)	1.000000	1.000	5	5	5 hit species in 5 species	5 hit proteins in 5 hit species from 5 total proteins in 5 total species	A.thaliana, C.elegans, D.melanogaster, D.rerio, H.sapiens
KOG2083 (4)	0.610577	0.619	3	4	3 hit species in 4 species	13 hit proteins in 3 hit species from 21 total proteins in 4 total species	C.elegans, D.rerio, H.sapiens
KOG1288 (6)	0.583333	0.571	4	6	4 hit species in 6 species	4 hit proteins in 4 hit species from 7 total proteins in 6 total species	C.elegans, D.rerio, H.sapiens, S.cerevisiae
KOG2082 (5)	0.542857	0.733	3	5	3 hit species in 5 species	11 hit proteins in 3 hit species from 15 total proteins in 5 total species	C.elegans, D.rerio, H.sapiens
KOG0224 (6)	0.466667	0.444	3	6	3 hit species in 6 species	8 hit proteins in 3 hit species from 18 total proteins in 6 total species	C.elegans, H.sapiens, S.cerevisiae
KOG0223 (6)	0.462838	0.197	4	6	4 hit species in 6 species	13 hit proteins in 4 hit species from 66 total proteins in 6 total species	A.thaliana, C.elegans, H.sapiens, S.cerevisiae
KOG4237 (4)	0.369048	0.278	3	4	3 hit species in 4 species	5 hit proteins in 3 hit species from 18 total proteins in 4 total species	C.elegans, D.rerio, H.sapiens
KOG0660 (7)	0.107143	0.065	3	7	3 hit species in 7 species	4 hit proteins in 3 hit species from 62 total proteins in 7 total species	D.melanogaster, D.rerio, S.cerevisiae
KOG3913 (4)	0.084803	0.088	3	4	3 hit species in 4 species	5 hit proteins in 3 hit species from 57 total proteins in 4 total species	D.melanogaster, D.rerio, H.sapiens
Group	Average H/M	Total H/M	Hit species	Total species	Hit species in total species	Description	List of Hit species

Showing 1 to 9 of 9 entries Previous 1 Next

Download this table in CSV file.

Figure 2.3-8: GO Regulation of cell size: Current annotations

The 9 orthologous groups that have at least three proteins annotated in at least three species are shown.

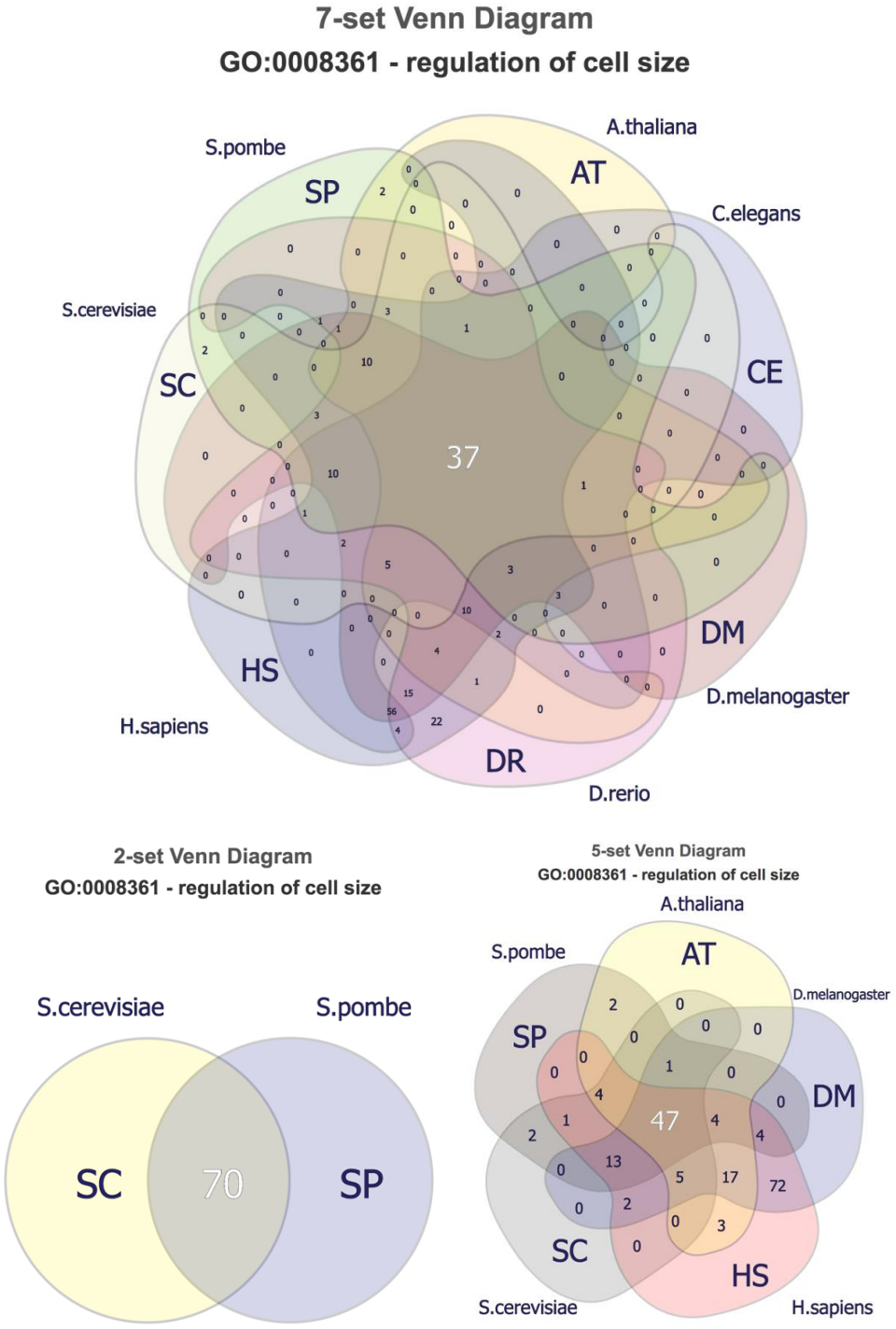


Figure 2.3-9: GO Regulation of cell size: Venn Diagrams

The diagram shows that there are more common protein groups in phylogenetically closer species than in more distant species. Between 2 species there are 70 common protein groups, while among 7 species there are only 37 groups of proteins.

Hit species threshold is 1.

Table 2.3-8: Number of GeneOntology annotations for the "GO:0008361 - regulation of cell size" GO term

The table shows that the Gene Ontology database is updated from time to time and sometimes annotations are not only corrected, but also erased from the database.

	GO:0008361 - regulation of cell size		
	Nr of proteins, 10th August 2017	eggNOG groups, 10th August 2017	Nr of proteins, 24th March 2018
A.thaliana	17	11	35
C.elegans	99	27	105
D.melanogaster	39	20	160
D.rerio	139	28	88
H.sapiens	204	91	209
S.cerevisiae	34	20	48
S.pombe	10	10	15

2.4 Conclusions

In this Chapter, I explained how I built two bioinformatics tools. The first directed to provide predictions of proteins that regulate cell size, while the second is a generalised tool to identify the conserved core associated with a GO function and to give novel functional annotations for the Gene Ontology database.

2.4.1 Ortholog Finder Tool results

I created the **Ortholog Finder Tool** that collects cell size related information from five model organisms, in parallel with the collection of their orthology relations. The tool can show functional orthological relations in line with known cell size data information. I have used the tool to predict functional orthology groups for subsequent wet laboratory experiments. With the help of these results, I was able to select gene orthology groups for experimental tests in *Arabidopsis* and also with Human T cells.

2.4.1.1 Results of experiments on selected proteins

I carried out experiments to determine whether the selected proteins have a role in regulating cell size. These experiments were based on the hypothesis that ablating a protein by gene knock-out in *A. thalina* or reducing the induction of the protein in human T cells during entry into the cell cycle from quiescence by transfecting siRNA would affect cell size. Thus, I created an analytical method, which I named the 'one-to-one method', which means that in all investigated species (five-species at the time: *A. thaliana, D. melanogaster, H. sapiens, S. cerevisiae* and *S. pombe*) the selected orthologous groups have to contain only a single copy of the ortholog gene. I found 17 orthology groups that could be further investigated experimentally (see the list of these groups in Table 2.3-2). I analysed this list manually and after reading the literature I selected TCTP/TPT1, TP53RK, VPS18/PEP3 TNPO3, and VBP1/PFD3 orthologs for further experiments.

TCTP/TPT1 orthologs are highly conserved, multifunctional proteins that are involved in many fundamental biological processes and disorders in human and other species (Bommer, 2017). TCTP/TPT1 orthologs have been reported to be involved in

mammary carcinoma (Gross et al., 1989), histamine release (MacDonald et al., 1987, 1995) and it can serve as an anti-apoptotic protein (Li et al., 2001). While the orthologs of the other selected protein, TP53RK/Bud32 are known to regulate p53 by phosphorylation on Serine 15 (Miyoshi et al., 2003). This phosphorylation of S15 impairs the binding of MDM2 and leads to the accumulation and activation of p53 (Shieh et al., 1997; Tibbetts et al., 1999). Although these two proteins have a wide range of characterised phenotypic effects in various organisms, they have never been associated with cell size regulation in Humans. The following chapters (Chapters 3 and 4) contain details of experiments designed to investigate whether TCTP and TP53RK have roles in regulating cell size in human T cells and In *Arabidopsis*. Briefly, the TCTP siRNA knock-down experiments on human T cells confirmed its role in cell size regulation. In contrast, reducing the levels of TP53RK did not. The protein encoded by the *Tctp1* ortholog in Arabidopsis was already known to have a role in regulating cell size (Berkowitz et al., 2008). However, as described in Chapter 4 I was not able to reproduce these results.

The VPS18/PEP3 protein is known to control the fusion of endomembranes (Hunter et al., 2017). VBP1/PFD3 a component of the prefoldin complex (Siegert et al., 2000), while TNPO3 is known to be responsible for transporting serine/arginine-rich proteins (Allemand et al., 2002; Kataoka et al., 1999). Although of the roles of all these proteins in several biological processes were investigated, they have never been reported to be involved in the regulation of cell size either in Human or *Arabidopsis*. In Chapter 4, I show that the *PFD3* knock-out mutants have a cell size defect in *Arabidopsis*. Since this protein is a member of a conserved protein complex, I decided to expand my analysis to other members of this complex.

Following the encouraging results on the cell size perturbing phenotype of the *PFD3* mutant in *Arabidopsis*, I further investigated the other members of the prefoldin complex (PFD1, PFD2, PFD4, PFD5 and PFD6). Prefoldin complex members are known to regulate the correct folding of proteins with molecular chaperonins (Vainberg et al.,

1998). Furthermore, some members have other functions, for example VBP1/PFD3 is a member of the ubiquitin protease system, which controls protein degradation (Mousnier et al., 2007; Tsuchiya et al., 1996). The PFD6 protein is known to be required for normal microtubule dynamics (Gu et al., 2008). Perturbation of *Pfd6* also affects plant size as these mutant plants were shown to be smaller than a control group (Gu et al., 2008). Other prefoldins have not been associated with any cell or organ size related effects in human, while *Pfd3* and *Pfd5* have been described to cause a specific size difference in *Arabidopsis* (Rodríguez-Milla & Salinas, 2009).

2.4.1.2 Possible further candidates

As the list of orthologous groups for possible targets contained 17 groups (Table 2.3-2), I could not Investigate 12 of them in my wet-laboratory experiments. If time had allowed, I would have selected and investigated a few of them. For example, the SWC4/DMAP1 orthologous group, which is a conserved member of the NuA4 histone acetyltransferase complex, that has orthology from yeast to humans (Doyon et al., 2004). DMAP1 has been reported to regulate ATM kinase activity, a survival function of the cells to halt the cell cycle while specific DNA damage is being repaired (Penicud & Behrens, 2014). These could affect cell size as controlled degradation of specific cyclins is Important In controlling progression through specific cell cycle phases and may effect cell size, which has been shown in budding yeast (Martínez-Láinez et al., 2018). In *Arabidopsis*, SWC4 as a part of the SWR1 protein complex contributes to plant development and physiology (Bieluszewski et al., 2015) and defects or deletion of the gene might affect the size changes normally associated with cellular differentiation.

2.4.2 The Gene Ontology Orthology Tool

I developed the **Gene Ontology Orthology Tool** that collects the current Gene Ontology functional annotations and shows them in line with functional orthologous groups exported from the eggNOG database (Huerta-Cepas et al., 2016; Jensen et al., 2008). The tool can identify the conserved core molecules of a given Gene Ontology term. These conserved core molecules seem to be crucial for the given pathway and may reveal novel insights of a regulatory mechanisms behind. In addition, the tool provides suggestions for novel protein functional annotations to a given GO term. These suggestions are purely based on my computational analysis, explained earlier in Section *2.2* and require experimental investigation.

2.4.2.1 Extension of GO annotations

The Gene Ontology (GO) database is not complete. Research groups have already created tools and approaches to expand and correct current annotations (The Gene Ontology Consortium, 2017). GO contains terms as a universal concept of relation to gene function and these terms are always expanded, revised and checked. There are tools like **TermGenie** (Dietze et al., 2014) that aims to create new classes *via* a web-based interface or a GitHub tracker (`https://github.com/geneontology/go-ontology/issues`) where issues can be posted for the structure of the ontology or new relationship types. There are also discussions taking place between the manual curators and members of the consortium about further possible annotations (The Gene Ontology Consortium, 2017).

As GO is a part of Open Biomedical Ontology (OBO) classes and using the OWL (Web Ontology Language), there are also regular updates from other ontologies. These updates include entries from **PO** (Plant anatomy) (Jaiswal et al., 2005), **ChEBI** (chemicals) (Degtyarenko et al., 2008) and **PATO** (qualities/descriptors) (Gkoutos et al., 2005).

2.4.2.2 Novelty of GOOT

As I described in section 2.1.2.4.4 , there are also approaches taking place in the GO community to expand GO based on orthological relations and homology (see Section 2.1.2.4.2). To find similar functional orthologs the method needs to employ homology similarity search by sequence, domain or protein structure (Pearson, 2013). In the current approach there are approaches to expand GO based on homology, such as InterPro2GO (Mitchell et al., 2015). InterPro2GO is an approach to integrate the functional knowledge of protein families, domains and sites which are combined from a number of different protein signature databases, such as Gene3D, Panther, Pfam, PRINTS, ProSite, SMART, SUPERFAMILY and TIGRFAM, then annotate them into GO (Mitchell et al., 2015). These methods describe the same protein family or domain into a single unique entry then applies cross referencing to include other proteins with similar protein sequence. If the InterPro finds similar proteins with a similar conserved function, it suggests a mapping between the InterPro entry and a GO database (Mitchell et al., 2015). There are other tools such as PAINT (Gaudet et al., 2011), which predict novel GO annotation based on phylogenetical relations such as sequence homology.

My method is a little different, since none of the other approaches currently uses a functional orthology database, such as the eggNOG database (Huerta-Cepas et al., 2016; Jensen et al., 2008) as a comprehensive source of functional annotation. eggNOG is a functional orthology database that groups proteins of 2031 eukaryotic and prokaryotic organisms into thousands of functional orthology groups (Huerta-Cepas et al., 2016). Using eggNOG made it easier to use the data, however it constrains my predictions to the current annotations in the eggNOG database. Moreover, current methods stick to the same taxonomic clade and they do not analyse similarities between mammalian and plant, whereas my tool is able to analyse proteins in Human and *Arabidopsis*. I used seven distinct, well-studied model organisms (*A. thaliana, C. elegans, D. melanogaster, D. rerio, H. sapiens, S. cerevisiae* and *S. pombe*) to find novel annotations for proteins and genes. It is for further research how far this can be expanded to other, less well-studied species.

2.4.2.3 Predictions of GOOT

2.4.2.3.1 Predicted novel annotations to GO

The tool predicted new GO annotations for 13 proteins in 4 species for the "*GO:0051726 - regulation of cell cycle*" GO term, where the total H/M value is 0.8 or above and hit species is one less than the total species. These proteins usually hold an uncharacterized name, or a name related to a hypothesised function (as for CyclinH in *Drosophila*), but they have not been annotated to the Gene Ontology. This information means that these proteins have been confirmed to exist and the genes encoding these proteins have been sequenced but have not been annotated to a function. The tool described here could extend what Is known by giving suggestions for Gene Ontology annotation predictions. However, there could be a bias in the prediction, because my main source of data is the eggNOG database. If a group of proteins is in the database, the tool can make predictions based on that protein, but if it is not it cannot make any predictions. My predictions can be varied by changing the H/M values which affects the probability of the result. I showed that this method could work in some cases such as "*GO:0051726 - regulation of cell cycle*". However, in other cases the currently available GO annotation data was not enough to identify a conserved core and give predictions of missing GO terms. I showed an example for the "*GO:0008361 - regulation of cell size*" GO term, where the situation was similar.

Chapter 3

Human T-Cells

3.1 Introduction

3.1.1 Cell size attributes in human cells

There are an at least 3.72×10^{13} cells in a human body (Bianconi et al., 2013) which have been classified as more than 411 different cell types (Vickaryous & Hall, 2006). Although these are large numbers and cells of different types are of a great variety of sizes, cells of the same type are similar in size. In addition, cells of a particular type maintain a specific range of sizes in a given tissue (Guertin & Sabatini, 2006). Therefore, there is a mechanism for sensing and maintaining a narrow range of cell sizes in human tissues. Three different parameters are controlled, namely the number of cell organelles, the total number of cells and their size (Marshall, 2016).

Organelles such as the nucleus (Jorgensen et al., 2007), nucleolus (Neumann & Nurse, 2007) and vacuoles (Berciano et al., 2007) scale with cells size (Chan & Marshall, 2012) and can affect the overall cell size. At least 5 different methods are known to control the size of organelles, which have been investigated by the use of organelle-specific reported molecules. To give some examples these are: (1) size dependent accumulation, if a reporter molecule binds to an organelle with which it has an affinity; (2) concentration gradient, a molecule in a certain part of the organelle creates a gradient depending on the size of the structure; (3) occupancy time, if a reporter molecule is delivered to a certain part of the organelle; (4) conformational change of a molecule that is dependent on the organelle and (5) structural scaffold (Chan & Marshall, 2012). The precise control of organelle size and number is important especially in Lymphocytes, but the overall cell size is mainly due to the size of the nucleus (Huber & Gerace, 2007).

Cell size regulation is dependent on cell growth, which is the mass accumulation of matter, and cell proliferation (Lloyd, 2013). Coordination between cell growth and cell proliferation is a key factor that sustains cells size through successive cell divisions (Conlon & Raff, 1999). The size of a given cell type is restored in dividing cells after each cell division and is maintained in quiescent cells. Cell growth and progression

through the cell cycle are coordinated in normal cells. However, these processes can be separated in mammalian cells with growth signals such as mTOR and PI3K (Fingar et al., 2002) and in human T-Lymphocytes (T-Cells) by peturbation of the regulatory "commitment point" (Lea, Orr, et al., 2003) or by siRNA experiments to prevent the induction of specific proteins normally caused by CD3/CD28 stimulation (Orr et al., 2012).

To test and analyse cell size changes in human cells, I picked T-Cells as a model. Specifically, quiescent, non-activated T-Cells which increase in size, enter the cell cycle and become functionally active in response to stimulation *via* CD3/CD28.

3.1.2 Human T-Lymphocytes

In my work, I have done experiments with non-activated human T-Cells isolated from peripheral blood. These are predominantly memory T-Cells and are $CD4^+$ and $CD8^+$ (see later for explanation). I will describe what is known about these cell populations in general in the sub-sections below.

3.1.2.1 T-Cells

T cells, also called *T-Lymphocytes*, are a type of white blood cell. These cells circulate in the bloodstream and they play a fundamental role in cell-mediated immunity. There are two major types of lymphocyte in a human body that determine the specificity of the immune response to antigens, namely T- and B-cells (Provan et al., 2004). T-Cell is a general term and many different subtypes of T-Cells are known, which have different Immune functions (Gerriets & Rathmell, 2012). T-Lymphocytes detect foreign substances, including infectious microorganisms as antigens and generate a specific immune response that can eliminate the threat from the body. T-Cells recognise non-self-molecules such as exogenously encoded pathogens upon ligation of clonotypic T cell receptors (TCR) by major histocompatibility complexes (MHCs) on antigen presenting cells (APCs) (Brownlie & Zamoyska, 2013). These TCR–peptide–MHC ligations induce an immune response against the pathogen by activation of effector molecules and T-Cell proliferation. The cells become effector T-Cells, and some become memory T-Cells (see details below) (Brownlie & Zamoyska, 2013).

3.1.2.2 Cell Size of T-Cells

White blood cells vary widely in shape and size when viewed by light microscopy, which is particularly evident after May-Grünwald-Giemsa (MGG) staining (see Section 3.2.12) and the population of lymphocytes is easily distinguishable because of cellular features. Most lymphocytes in the peripheral blood are non-activated and quiescent. These cells are oval or spherical with basophilic cytoplasm, a large spherical nucleus that fills up to 90% of the cell and contains condensed dark chromatin (see Figure 3.1-1). The usual cell size of normal, healthy lymphocytes is between 6.8 – 7.8 µm (Kuse et al., 1985) and the cells double in size after stimulation *via* CD3/CD28 (see below).

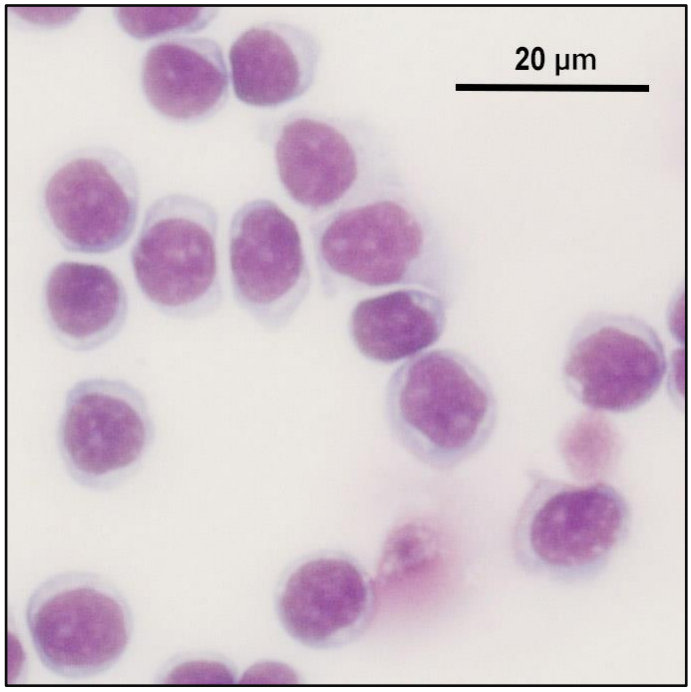


Figure 3.1-1: Isolated, unstimulated, quiescent T-Lymphocytes under a light microscope

This image is of quiescent, quiescent (G_0) T-Lymphocytes. These cells are oval or spherical with a small annulus of cytoplasm surrounding a large spherical nucleus containing condensed dark chromatin.

Non-activated, quiescent T-Cells were isolated from human peripheral blood as described in Section 3.2.4 , immobilised onto a microscope slide by Cytocentrifugation and stained with MGG dye as described in Section 3.2.12 .

I captured this image of quiescent T-Cells using light microscopy.

3.1.2.3 Origin of T-Cells

T-Cells originate from haematopoietic stem cells in the bone marrow, but they mature in the thymus, which gives them the name *T*-Cells (Brownlie & Zamoyska, 2013; Schwarz & Bhandoola, 2006). If the thymus is absent, low numbers of T-Cells are produced (Boyd et al., 1993). See Figure 3.1-2 on how T-Cells are produced from haematopoietic stem cells.

Haemopoietic progenitor cells are kept in the bone marrow by interactions of cell surface receptors such as SDF-1 and

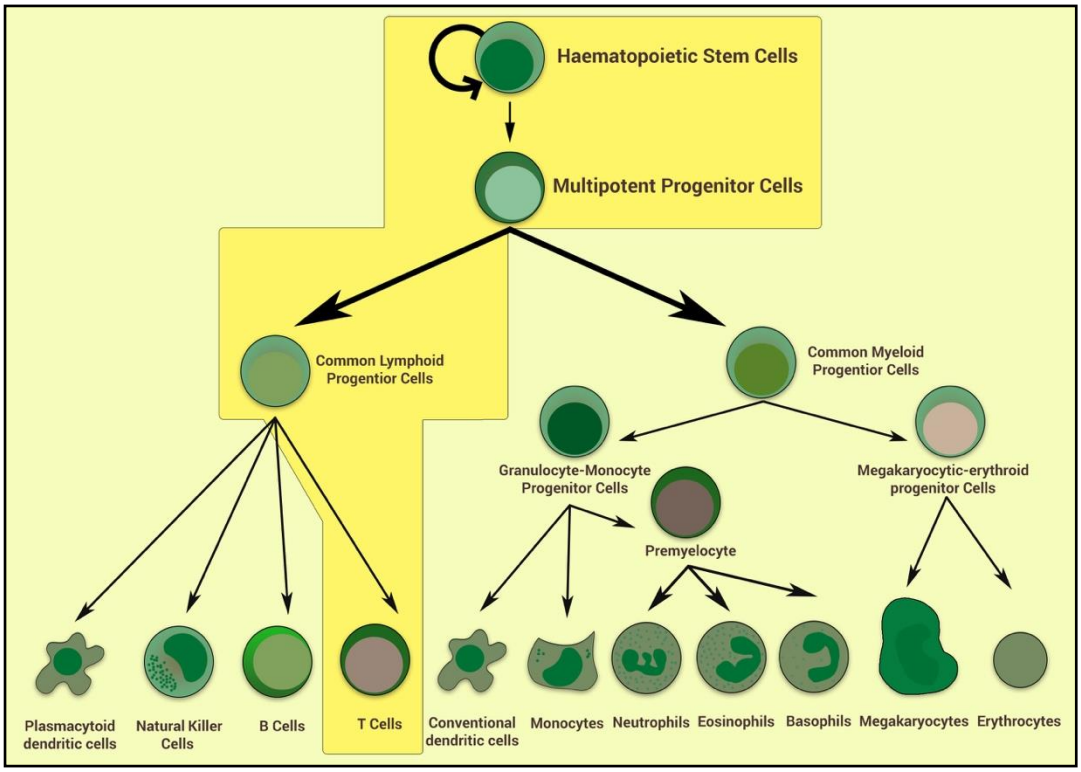


Figure 3.1-2. Representation of haematopoiesis that illustrates the main lineage commitment steps.

Hematopoietic Stem Cells are at the top of the figure. When these cells divide, *Multipotent Progenitor Cells* (MPC) are produced from one of the progenies and the other remains as a stem cell. The MPCs give rise to common lymphoid and common myeloid progenitor cells.

Lymphoid cells: There are four mature cell types, which are produced through a step-wise process of lineage commitment, *Plasmacytoid dendritic cells*, *Natural Killer cells*, *B-Cells* and *T-Cells*.

Myeloid cells: There are two types of progenitor cell: Granulocyte-Monocyte Progenitor Cells and Megakaryocytic-erythroid Progenitor Cells. These cells mature into seven different cell types through commitment steps to produce *Conventional dendritic cells, Monocytes*, *Neutrophils*, *Eosinophils*, *Basophils*, *Megakaryocytes* and *Erythrocytes*.

The figure is based on Figure 1 from Famili et al., 2017.

3.1.2.3.1 Subtypes of T-Cells

Several subsets of T-Cells exist, which all have distinct functions in the body. The majority of human T-Cells have α and β chains that make up their TCR (αβ T-Cells) (Krogsgaard & Davis, 2005). Several subsets of αβ T-Cells exist that have functions, such as effector, helper and cytotoxic T-Cells. These T-Cells are distinguishable by the fact that they express different co-receptors and produce various cytokines (see Figure 3.1-3 for details). Subsets of αβ T-Cells can be categorised into at least three different groups based on their function and co-receptors. The most common subsets of αβ T-Cells are distinguished by CD4 and CD8 co-receptors. These cells have a TCR complex that consists of 3 different proteins: (1) the αβ TCR (2) linking co-receptor molecule CD4 or CD8 and (3) specific MHC molecules such as MHC class I for CD8 and MHC class II for CD4 on APC cells (Davis et al., 1998). In humans, there are three polymorphic class I genes (*HLA-A, B-* and *-C*) and four class II gene pairs (*HLA-DM, -DP, -DQ* and *-DR*) (Ellis & Ballingall, 1999; Hughes, 1995). Lymphocytes make up 70-90% of peripheral blood mononuclear cells (PBMC), of which 70-85% are T-Cells. There is approximately a 2:1 ratio of $CD4^+$ and $CD8^+$ T-Cells. These are the predominant T-Cell populations that I isolate for the work described in this chapter (Lea, Orr, et al., 2003).

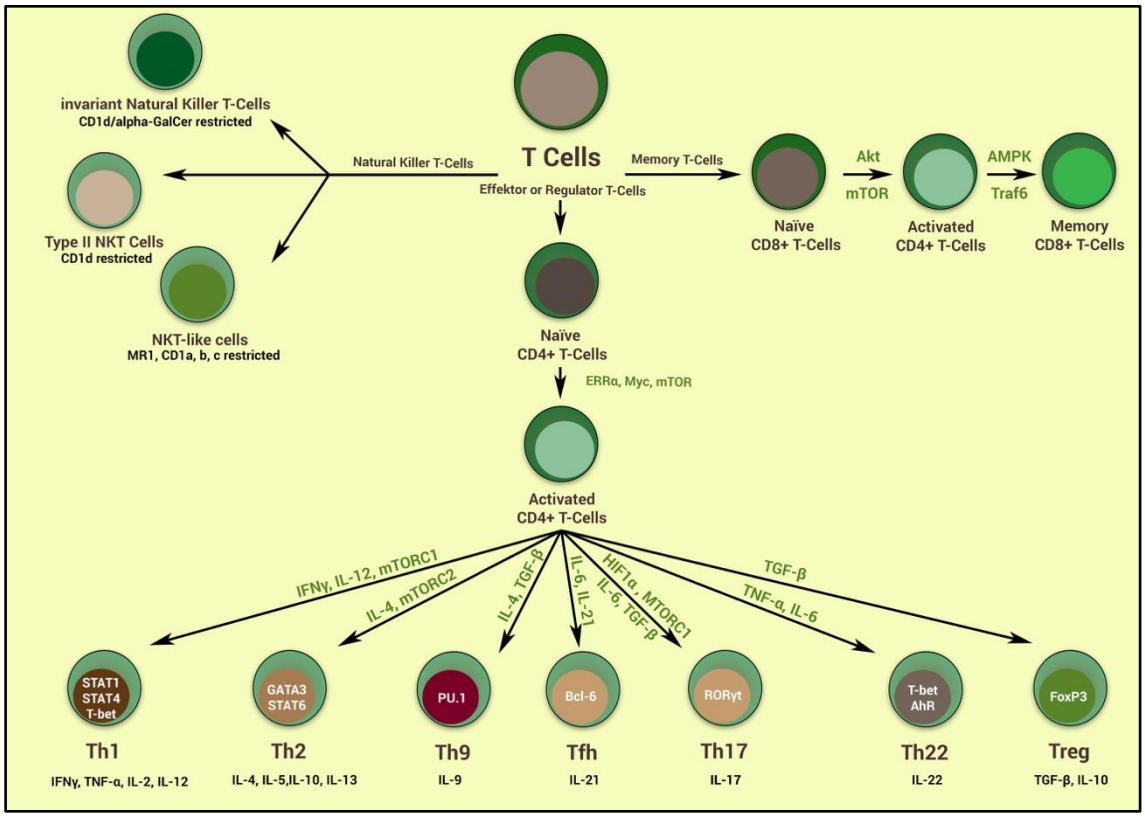


Figure 3.1-3. Different Subtypes of T-Cells.

There are two major groups of T-Cells. The most common group is **CD4⁺ T-Cells.** Naïve T-Cells become activated by antigen presentation. Responses to specific interleukins (IL), interferons (IFN) and other cytokines such as Transforming growth factor beta (TGF-β), Tumour necrosis factor alpha (TNF-α) and Hypoxia-inducible factor 1 alpha (HIF1α) cause differentiation to seven subtypes, namely Th1, Th2, Th9, Th17, Th22, Tfh and Treg. These cells are effector T-Cells, and all of them produce the cytokines indicated.

The second **group are CD8⁺ T-Cells**, which consist of naïve and memory cells. These cells are generated to combat infections.

The third and much smaller group is **Natural Killer T-Cells**, that is as heterogeneous group shares similar properties with Natural Killer Cells. These cells can be divided into three different subgroups.

The image is based on Figure 2 from Gerriets & Rathmell, 2012, Figure 1.1 from Sun & Zhang, 2014 book and Figure 1 from Simoni et al., 2013.

3.1.2.4 Activation and antigen discrimination

Most primary T-Cells in the human circulatory system are in an unstimulated, quiescent (G_0) state (see below). These cells remain in this quiescent state until they receive a specific mitogenic stimulus *via* their TCR. If a specific stimulus is received, T cells respond by activating effector functions, and they enter the growth cycle and cell cycle, as described below.

3.1.2.4.1 The Quiescent state

Quiescent T-Cells are in a non-proliferative cell cycle phase termed G_0 (Takada & Jameson, 2009; Thomas, 2004; Yang & Chi, 2018). These quiescent T-Cells are small in size relative to proliferating T cells with low protein content, 2n DNA content and low metabolic activity (Thomas, 2004). Quiescent T-Cells can enter the cell cycle and divide when they receive a mitogenic stimulus, such as *via* the TCR and co-receptors such as CD28 (see below). Quiescence is different from senescence since senescent cells have lost their ability to respond to mitogenic stimuli and to divide (Campisi, 1996; Terzi et al., 2016). The quiescent state is not just a dormant state with limited activity, it is an actively controlled state in which cells are still capable of responding to stimulation and is necessary for T-Cell homeostasis (Hamilton & Jameson, 2012; Yang & Chi, 2012). T-Cells are thought to be kept in G_0 by proteins, such as the hypo-phosphorylation of the retinoblastoma protein family, namely pRB and p130 (reviewed in Thomas, 2004). Many proteins that are required for progression through the cell cycle, such as minichromosome maintenance proteins (MCM2 to MCM7) and Cdc6 (Williams, Shohet, et al., 1997; Yan et al., 1998), DNA polymerase α (Nakamura et al., 1984), ribonucleotide reductase (Mann et al., 1988) and thymidine kinase (Rittling et al., 1986; Wintersberger et al., 1992) are not present in quiescent T-Cells. Therefore, these cells do not contain many of the proteins necessary for cell proliferation (Lea, Orr, et al., 2003; Orr et al., 2010). These proteins are synthesised *de novo* when T cells are stimulated to enter the cell cycle by ligation of CD3 and CD28 (*ibid*). Peripheral blood T-Cells produce energy in the quiescent state by catabolic metabolism, which is insufficient to support the increase in size that occurs in response to stimulation *via* CD3/CD28 (Rathmell et al., 2000). When T cells are stimulated by CD3/CD28

ligation, metabolic reprogramming occurs from oxidative phosphorylation (OXPHOS) to glycolysis (Allison et al., 2017; Buck et al., 2015, 2017; Fox et al., 2005; MacIver et al., 2013; van der Windt & Pearce, 2012). This is associated with increases in the levels of nutrient transporters (Orr et al., 2012), including the glucose transporter Glut1 which translocate to the cell surface (Wieman et al., 2007).

3.1.2.4.2 Activating T-Cells

T-Cell activation is initiated by APCs, which have MHCs bound to a peptide antigen on their surfaces, as described above. APCs can be dendritic cells, macrophages or B-Cells. Immune response to foreign antigens *via* APCs activate heterodimeric αβ TCR (Krogsgaard & Davis, 2005), described above. TCR αβ heterodimer receptors are non-covalently associated with three forms of the CD3 subunit dimers: CD3εγ, CD3εδ heterodimers and CD3ζζ homodimers (Call & Wucherpfennig, 2007; Ngoenkam et al., 2017; Weiss, 1993). A specific amino acid sequence, called immunoreceptor tyrosine-based activation motif (ITAM) can be found on the cytoplasmic side of CD3 subunits: CD3ε, CD3γ and CD3δ contain one ITAM, while CD3ζ contains 3 ITAMs (Brownlie & Zamoyska, 2013). Activation of the TCR-CD3 complex by antigen binding leads to intracellular signal transduction that initiates cell proliferation, increase in cell size and activation programmes (Brownlie & Zamoyska, 2013).

3.1.2.4.2.1 CD3 / CD28 co-stimulation

Stimulation *via* the TCR-CD3 receptor is not enough to trigger a full immune response, and in the absence of a second stimulus *via* a co-stimulatory receptor an abortive immune response is the result (Jenkins et al., 1990). In this case, the T-Cell is unable to produce interleukins and is not able to proliferate (Porciello & Tuosto, 2016; Schwartz, 2003). There are two major groups of co-stimulatory proteins on the surfaces of APCs. First is the ligation of CD28 by B7.1/CD80 or B7.2/CD86 (Freeman et al., 1993; Gross et al., 1990), which both result in transcriptional activation of genes required for cell cycle entry and functional activation (Acuto & Michel, 2003). Gene transcription is activated *via* transcription factors such as NF-κB, PU.1, E2F and others that are

necessary for the proliferation and activation programs (Li et al., 2005; Rothenberg, 2014). Stimulation of CD28 receptors by either B7.1/CD80 or B7.2/CD86 results in increased expression of the interleukin 2 receptor (IL2R) on the cell surface (Shahinian et al., 1993), production of interleukin 2 (IL2) (Lucas et al., 1995; Reichert et al., 2001) and entry into the cell cycle (Bonnevier & Mueller, 2002). Specifically, CD28 receptor ligation upregulates the expression of D-type cyclins and causes a reduction of cyclin-dependent kinase inhibitors (CKI), such as $p27^{Kip1}$ that enables cycle entry to be initiated from quiescence (Appleman et al., 2002). Cyclins control progression through the cell cycle as well as other cellular processes, such as transcription by initiating the activation of Cyclin-dependent kinases (Cdk) (Hochegger et al., 2008). CD28 co-stimulation also activates the PI3K/mTOR signalling pathway (Mamane et al., 2006). mTOR is a major signalling pathway that leads to an increase in cell size, as described in Chapter section 1.3.3.1 Also, CD28 co-stimulation causes cytoskeletal rearrangement, resulting in a change of cell shape from small round cells (Figure 3.1-1) to larger cells with irregular shapes (Boomer & Green, 2010).

The consequences of stimulation *via* CD3/CD28 are that T-Cells enter the cell cycle, become larger, *i.e.* enter the growth cycle that controls cell size, and induce many effector molecules that are responsible for T-cell functions.

3.1.2.4.3 The Mitotic Cell Cycle

The mitotic cell cycle can be divided into two separate gap phases (namely G_1 and G_2 phases), which are separated by two synthesis phases (see Figure 3.1-4). Between G_1 and G_2 phases, there is a cell cycle phase during which DNA is synthesized (called S-phase), while cell division/cytokinesis occurs between G_2 and G_1 (called M-phase) (Thomas, 2004). The transitions between phases are coordinated by members of the Cdk family, which are activated by cyclins, as their name suggests (Hochegger et al., 2008). Non-stimulated, quiescent cells are in the G_0 phase described above, and upon activation, they enter the G_1 phase of the cell cycle (reviewed in Thomas, 2004).

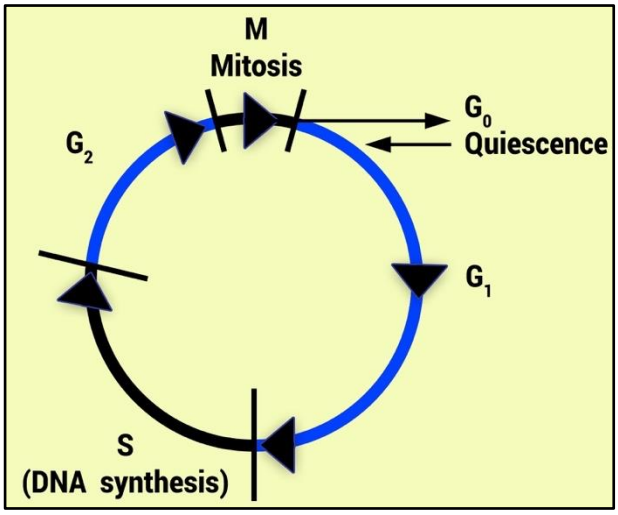


Figure 3.1-4: The main phases of the cell cycle

A mitotic cell cycle consists of four phases: mitosis (M phase) and DNA synthesis (S phase) that are separated by two gap phases called G_1 and G_2. Cells can exit the G_1 phase and enter a non-proliferative, quiescent state called G_0.

The figure is based on Figure 8.2 from Thomas (2004).

3.1.2.4.3.1 The Commitment Point during Cell Cycle entry from quiescence

In human cells there exist numerous control points that regulate progression through cell cycle phases (Charollais et al., 1994). The Thomas laboratory identified the "*commitment point*" as the point during the transition from G_0 to G_1 phase after which continued stimulation of CD3 and CD28 are no longer needed for T-Cells to progress into the cell cycle (Lea, Orr, et al., 2003). This study identified that T-Cells require 3-5 hours of continuous CD3/CD28 stimulus to commit to entering the cell cycle. If the stimulus is stopped earlier than this, cells remain quiescent. This G_0 - G_1 commitment point requires the activation of CDK4/6 – Cyclin D since cell cycle entry is inhibited by adding TAT-$p16^{INK4A}$, which enters the cells and inhibits both CDK4 and CDK6 activity. Transition through the commitment point is also necessary for T cells to enter the growth cycle and increase in size, but transition through the commitment point is not required for induction of effector functions. This study showed that entry into the cell cycle and the growth cycle could be uncoupled from the expression of T-Cell early effector functions (Lea, Orr, et al., 2003). A subsequent study from the laboratory showed that entry into the cell cycle from quiescence and entry into the growth cycle could also be uncoupled (Orr et al., 2012).

3.1.2.4.3.2 Checkpoints during the G_1 phase

There exists another control point during the G_1 phase called the restriction point. This restriction point (R-point) is the point beyond which cells no longer need growth factors to enter the S phase of the cell cycle. Although the R-point was identified using fibroblasts (Pardee, 1974) an analogous control point may exist in other cell types, such as T-Cells (Cantrell & Smith, 1983). The restriction point in T-Cells involves stimulation *via* the IL-2 receptor. IL-2 withdrawal causes T-Cells to exit the cell cycle and arrest in mid/late G_1. Re-stimulation with IL-2 causes re-entry into S-phase, which takes a shorter time than the transition into S-phase from quiescence (Cantrell & Smith, 1984; Cantrell et al., 1988). However, in T-Cells, there is an activation-induced cell death program that protects against inappropriate, repeated T-Cell proliferation by triggering apoptosis (Rodríguez-Tarduchy & López-Rivas, 1989; Shi et al., 1990). IL-2 control is executed in part by causing degradation of the $p27^{kip1}$ cyclin-dependent

kinase inhibitor (CKI) (Boonen et al., 1999; Moore et al., 2001). In the absence of the IL-2 signal, $p27^{kip1}$ functions as an inhibitor of cell cycle progression as it inhibits the activation of the cyclin-dependent kinase Cdk2-Cyclin E (Pavletich, 1999; Sherr, 1994).

Various cyclin–CDK complexes are involved in progression through G_1 (Hengstschläger et al., 1999) by phosphorylating proteins such as pRb. When pRb is in a hypophosphorylated state, it binds and inhibits E2F transcription factors (Weinberg, 1995). Upon phosphorylation by cdk4-Cyclin D and cdk2–Cyclin E complexes on sites including serine 807 and 811, it releases E2F transcription factors to activate the transcription of genes encoding proteins required for progression through the cell cycle and other cellular programmes (MacDonald & Dick, 2012; Thomas, 2004; Weinberg, 1995). pRb is one of a family of three proteins, the others being RBL1 (p107) and RBL2 (p130). Both of these proteins control the progression of cells into the cell cycle by inhibiting E2F activity (Cobrinik, 2005) and there is evidence that the main complex in quiescent haemopoietic cells, including T-Cells is p130-E2F4-DP1 (Thomas et al., 1998; Williams, Linch, et al., 1997).

3.1.2.4.3.3 The function of Cyclins, CDKs and CKIs

The Cyclin family are a conserved group of proteins which are fundamentally involved in regulating the cell cycle. Cyclins were originally discovered as proteins that are synthesised and degraded cyclically as cells progress through the cell cycle (Evans et al., 1983). Cyclins are subunits of CDK holoenzyme[5] (cyclin-dependent kinase) complexes which regulate cell cycle checkpoints by phosphorylating specific target proteins on Serines and Threonines (Reviewed in Goodman, 2008). To date, more than 20 cyclins and cdks have been discovered in higher eukaryotic cells, many with unknown functions (Goodman, 2008; Kalra et al., 2017).

The discovery of cyclins and CDKs were major scientific breakthroughs that were celebrated by the Nobel prize in 2001 awarded to Tim Hunt, Paul Nurse and Leland

[5] Holoenzyme is a biochemically active molecular complex that is formed of an enzyme and a coenzyme.

Hartwell. Hartwell found genetic mutants of yeast that affect the cell cycle, which he termed cell division cycle (CDC) mutants. He identified *CDC28*, which affects progression through G_1 and G_2/M in *S. cerevisiae* (Hartwell et al., 1970). Later Paul Nurse identified the *cdc2* gene, a homologue that controls mitosis in another yeast, *S. pombe* (Nurse & Thuriaux, 1980). The regulatory mechanisms involving CDC28/Cdc2 are conserved in all eukaryotes (Lee & Nurse, 1987). Cdc2 (known as Cdk1 in higher eukaryotes) is the only cyclin-dependent kinase that is required in mice and *CDK1*$^{-/-}$ mice are not viable as no cell division takes place (reviewed in Malumbres & Barbacid, (2009)).

The literature currently categorizes the cyclins to two groups based on their involvement in regulating the cell cycle. There are G_1/S cyclins, such as the D-type cyclins which regulate CDK4/6 (Sherr & Roberts, 1999) and E-type cyclins which regulate CDK2 (Sherr & Roberts, 1995)(Ajchenbaum et al., 1993). The amount of Cyclin-D3 is increased significantly in T-Cells that have encountered a mitogen stimulus, which suggest that this cyclin may also have a role in responding to external stimuli that cause procession into the cell cycle from quiescence (Lea, Orr, et al., 2003). The second group are G_2/M cyclins, which include B-type cyclins which regulate CDK1 (Porter & Donoghue, 2003).

Cdk activity is under the control of cyclin-dependent kinase inhibitors (CKIs). Two major groups of CKIs can be categorised. The first group is INK4: $p16^{INK4A}$, $p15^{INK4B}$, $p18^{INK4c}$ and $p19^{INK4D}$, which are inhibitors of Cdk4 and Cdk6 (Jeffrey et al., 2000). The second group of CKIs are Cip/Kip proteins, and members of this group are $p21^{Cip1}$, $p27^{Kip1}$ and $p57^{Kip2}$. These enzymes are known to inhibit cdk2 complexes (Vidal & Koff, 2000). However, depending on the stoichiometry, $p21^{Cip1}$ can either stimulate or inhibit Cdk2-Cyclin E/A activity and CKIs also have roles in other cellular functions such as apoptosis, transcription and cell fate (Besson et al., 2008).

3.1.3 Regulation of the size of human cells

As I have stated above T-Cells are under rigorous regulation and cell size is controlled within a narrow range. The growth of T-Cells (Blastogenesis) is regulated by mTOR pathways both in the quiescent state *via* Tsc1 (Yang et al., 2011) and in an activated state (Fingar et al., 2002; Terada et al., 1995). Indeed, a study suggests that development of T-Cells can be tracked by cellular size, which is regulated by mTOR activity (Pollizzi et al., 2015).

3.1.3.1 Reported cell size regulators

There are many known cell size regulators of T-Cells, some of them are related to the mTOR pathways, while some of them are independent of mTOR. Two examples are the Cytotoxic T lymphocyte antigen-4 (CTLA-4), an inhibitor of T-Cell responses which regulates the size of $CD4^+$ T-Cells during an early phase of activation (Kuhns et al., 2000). Studies suggest that CTLA-4 is connected to the Akt/mTOR pathways (Karman et al., 2012). Another mTOR independent (Miluzio et al., 2016) example is eIF6 (Orr et al., 2012). The eIF6 protein regulates 60S ribosome biogenesis and my laboratory showed that reducing the induction of eIF6 in T-Cells during the G_0 to G_1 transition with siRNA results in cell cycle entry at a reduced cell size (Orr et al., 2012).

3.1.3.2 The quest for novel cell size regulators

Silencing or reducing the expression of a given gene can be achieved by using CRISPR-Cas9, ribozymes, DNAzymes, antisense oligodeoxynucleotides (ASO) or small interfering RNA (siRNA). The siRNA process is reviewed by Scherer and Rossi (2003) (Scherer & Rossi, 2003). To find novel cell size regulators, I followed an siRNA-mediated "knock-down" method that was published previously by our laboratory (Lea, Buggins, et al., 2003; Orr et al., 2010, 2012). As described above, quiescent T-Cells do not contain many proteins that are essential for progression through the cell cycle and other cellular processes, or these proteins are expressed at very low levels. CD3/CD28 stimulation induces the genes encoding these proteins that are required for progression through the cell cycle. The Thomas laboratory showed that 100% of quiescent T cells can be transfected with siRNA (Lea, Buggins, et al., 2003) and this method could be used to reduce the induction of specific mRNA and hence the protein encoded, which would normally occur in response to CD3/CD28 stimulation (Orr et al., 2010, 2012).

3.1.4 Predicted cell size regulator proteins

As I showed in the previous Chapter, a distinct group of proteins have been identified that are potentially capable of regulating the size of human cells (see in Chapter 2.3). I selected TCTP/TPT1, TP53RK, VBP1, VPS18 and TNPO3 as candidates for experiments with human T-Cells, together with the prefoldin complex protein **VBP1** (von Hippel-Lindau binding protein 1). The DNA replication protein **MCM7** is also included in the screen as a technical control for the effectiveness of siRNA transfection and siRNA-mediated reduction of the induction of Mcm7, which normally occurs in response to CD3/CD28 stimulation (Orr et al., 2010).

3.1.4.1 TCTP protein in human

Translationally Controlled Tumour Protein (TCTP) is a highly conserved, multifunctional protein that is involved in many fundamental biological processes and disorders either in human or other species (Bommer, 2017). The protein is located in the cytosol (Sanchez et al., 1997) and the nucleus (Cheng et al., 2012) of cells. It is encoded by the Tumour Protein, Translationally-Controlled 1 (*TPT1*) gene that is located on Chromosome 13 between 13q12 – q1413 (MacDonald et al., 1999).

3.1.4.1.1 The history of the TCTP protein

The TCTP protein was first identified more than 30 years ago as *Q23* in the serum-stimulated Swiss 3T3 mouse cell line (Thomas et al., 1981), then a year later in mouse sarcoma ascites cells as *P21* (Yenofsky et al., 1982). The mouse cDNA was cloned and sequenced in 1988 (Chitpatima et al., 1988). At the time, the function of the protein was unknown, and naming was inconsistent in the literature, merely based on its measured molecular weight. In human, the protein was first discovered in Ehrlich ascites tumour cells as *P23* (Benndorf et al., 1988). A year later the protein was named as "*Translationally Controlled Human Tumor Protein*" (Gross et al., 1989) based on the fact that the synthesis of the protein is under translational control in the ascites tumour cell line (Böhm et al., 1989) and the human cDNA was cloned from human mammary

carcinoma cells (Gross et al., 1989). TCTP is a protein that has a wide range of effects, including histamine release (named *Histamine-releasing factor (HRF)*) (MacDonald et al., 1987, 1995) and as an anti-apoptotic protein (*fortilin*) (Li et al., 2001). Since early 2000 the naming has been consistent in all species where the protein or its homologues are present (Bommer, 2017). TCTP is the protein name, and *TPT1* is the commonly used as gene name (Thiele et al., 2000), with the exception of yeast, where the gene symbol is *YKL056c,* and the protein name is TMA 19 (Fleischer et al., 2006).

3.1.4.1.2 Conserved protein

Studies show that the sequence of TCTP is evolutionarily conserved in eukaryotes (Hinojosa-Moya et al., 2008; Thayanithy & Venugopal, 2005; Yubero et al., 2009)(Hinojosa-Moya et al., 2008). The TCTP protein is expressed in all eukaryotic cells and tissues where it has been investigated (Bommer, 2017; Hinojosa-Moya et al., 2008; Thiele et al., 2000), although the level of expression is significantly lower in terminally differentiated tissues (Bommer & Thiele, 2004). TCTP seems to be an essential protein in mammalian early development, as *TPT1*$^{-/-}$ is embryonic lethal in mice (Chen et al., 2007; Koide et al., 2009; Susini et al., 2008)

3.1.4.1.3 The TPT1 gene, mRNA and 3D structure

The structure of the *TPT1* gene of 4000 nucleotides was first characterised in rabbit and has six exons, and five introns (Thiele et al., 1998) and the human *TPT1* gene is similar (Andree et al., 2006). The transcription of the gene results in two different mRNAs, both in rabbit (Thiele et al., 2000) and human cells (Andree et al., 2006), while the open reading frame of the mammalian mRNA encodes a protein of 172 amino acids (Thiele et al., 2000). The 3D structure of the yeast protein was first characterised, and this revealed similarity with the Mss4/Dss5 protein family (guanine nucleotide-free chaperones) (Thaw et al., 2001). The characterisation of the human structure revealed a folded core region in line with a long flexible loop (Feng et al., 2007; Susini et al., 2008). The 3D structure of the human protein has eleven β-strands with three α-helixes (Figure 3.1-5).

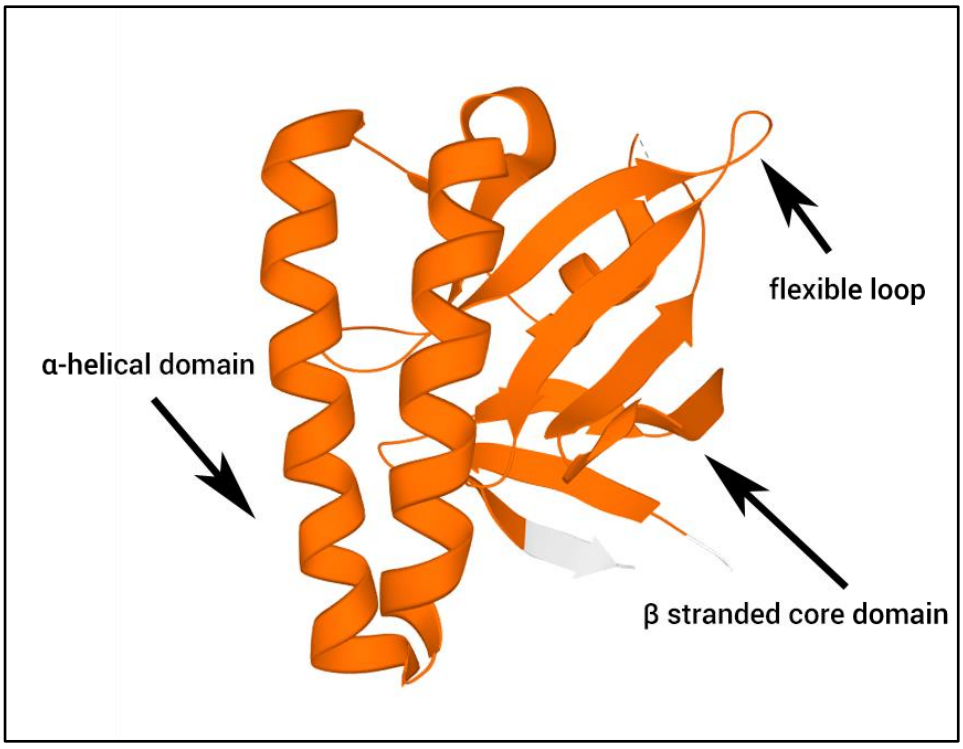


Figure 3.1-5: 3D structure of human TCTP protein

Two major secondary structure elements of the protein can be recognised: a β stranded core domain that consists of eleven β strands from three different β sheets, with three α-helixes, a more extended pair of them at the front and a shorter one at the back. These connect to a flexible loop region (thin lines). An α-helical hairpin connects the two α-helixes.

The Figure was taken with Protein Data Bank 3D Structure viewer using EMBL-EBI website (http://www.ebi.ac.uk/pdbe/entry/pdb/).

Source PDB id: 1yz1 (Susini et al., 2008)

3.1.4.1.4 Function of TCTP

3.1.4.1.4.1 Chaperonin functions of TCTP and ion homeostasis

TCTP functions as a heat shock protein with chaperonin like activity (Gnanasekar et al., 2009; Mak et al., 2007) and conserved members of the HSP70 family interact with TCTP (Li et al., 2016). The protein has a role in sensing and repairing DNA damage (Zhang et al., 2012). TCTP also binds Ca^{2+} and plays a substantial role in maintaining ion homeostasis (Haghighat & Ruben, 1992).

3.1.4.1.4.2 TCTP and Apoptosis

Another key function of TCTP is its anti-apoptotic activity (Bommer, 2017). Anti-apoptotic proteins from the Bcl2 family cooperatively interact with TCTP (Liu et al., 2005; Zhang et al., 2002) and TCTP is thought to prevent apoptosis by counteracting the pro-apoptotic activity of Bax (Susini et al., 2008). TCTP also prevents p53-mediated apoptosis by stimulating its degradation (Amson et al., 2012; Chen et al., 2011; Rho et al., 2011).

3.1.4.1.4.3 TCTP in mitosis and meiosis

The TCTP protein has been shown to locate to microtubules during the cell cycle (Gachet et al., 1999) and the protein plays a role in binding to the mitotic spindle poles (Bazile et al., 2009; Jaglarz et al., 2012). TCTP also has a function in stabilizing microtubules during meiosis (Tani et al., 2007).

3.1.4.1.4.4 TCTP and cell growth

It has been revealed that TCTP protein levels are transcriptionally regulated *via* PI3-Kinase / mTOR pathways in *Drosophila* (Hsu et al., 2007) and human cells (Dong et al., 2009). However this finding has been challenged (Wang et al., 2008). The protein reportedly has a direct effect on cell growth (Bommer et al., 2015). In the other hand, TCTP binds to one of the cytoplasmic domains of the Na, K-ATPase, that results in the activation of Src kinase that triggers the PI3K/Akt and Ras/Raf/ERK pathways (Jung et al., 2011).

3.1.4.1.5 TPT1 and cancer

3.1.4.1.5.1 TCTP role in cancer cells

Despite the name transitional tumour control protein, the finding that TCTP proteins are linked to cancer was made only late in 2009 (Telerman & Amson, 2009). TCTP protein levels are overexpressed in human tumours and cancer cells, while they have also been shown to downregulate tumour suppressor genes, such as p53 (Acunzo et al., 2014; Telerman & Amson, 2009). It has been shown that TCTP is also capable of promoting the transition from epithelial cells to mesenchymal cells, that is a fundamental step in invasive cancer (Bae et al., 2015). While a study also reported that reducing the expression of the TCTP protein inhibits the proliferation and invasion of LoVo colon cancer cells (Chu et al., 2011; Ma et al., 2010)

3.1.4.1.5.2 TCTP: Cell growth in cancer

It has been reported that TCTP actively induces proliferation and cell growth in early stages of colorectal cancer at the adenoma stage (Bommer et al., 2017; Xiao et al., 2016) and it is known to be upregulated through the PI3K/Akt pathway (Zhang et al., 2009).

3.1.4.1.5.3 TCTP: Mitosis in cancer

Mitotic functions of TCTP were investigated using hepatocellular carcinoma cells. The studies found that overexpression of TCTP occurs as a result of CHD1L oncogene activity (CHD1L is a chromatin helicase) and the number of defects during mitosis increased (Chan et al., 2012).

3.1.4.1.5.4 TCTP and chemoresistance

The connection between chemoresistance and the TCTP protein was studied in non-Hodgkin lymphomas (He et al., 2015). Results of work on breast cancer showed that phospho-TCTP levels are an indicator of the mitotic activity of cancer cells that are resistant to treatment (Lucibello et al., 2015).

3.1.4.1.6 TCTP and cell size

TCTP has been reported to bind to a member of the TOR pathway called Rheb[6] in *Drosophila*, which leads to the regulation of cell number, cell size and cell growth (Hsu et al., 2007), and cell development (Hong & Choi, 2013). Reports of experiments using diabetic mouse cells also stated that TCTP expression is elevated in glomerular cells, plus the reduction of cell size in podocytes can be seen after the reducing the expression of the TCTP protein (Kim et al., 2012).

The TCTP protein has many cellular functions, as described above and it is unclear whether TCTP has a direct effect on cell growth. It may regulate cell size indirectly since it is a chaperonin molecule, which is responsible for the correct folding of a number of cellular proteins (Gnanasekar et al., 2009; Mak et al., 2007) (see Section 3.3.4.2.2 below).

3.1.4.2 TP53RK protein in humans

TP53RK, also known as PRPK (p53 related protein kinase) is a highly conserved 20 kDa protein kinase (Miyoshi et al., 2003). The TP53RK protein was first characterised in budding yeast as YGR262c (Stocchetto et al., 1997), then later in human IL-2 activated cytotoxic T-Cells (Abe et al., 2001). The protein was shown to regulate p53 by phosphorylation on Serine 15 (Miyoshi et al., 2003). Phosphorylation of S15 impairs the binding of MDM2 and leads to the accumulation and activation of p53 (Shieh et al., 1997; Tibbetts et al., 1999).

3.1.4.2.1 p53 protein

The p53 protein is a well-known tumour suppressor protein (Levine, 1997). The expression levels of p53 are controlled by MDM2, an E3 ubiquitin ligase (Hu et al., 2012) that causes the degradation of p53 (Haupt et al., 1997; Honda et al., 1997). It is

[6] Rheb, Ras homolog enriched in brain, is a small GTP-binding protein that is the target of the mTORC1 complex and leads to a phosphorylation cascade that results in cell growth and proliferation (Mazhab-Jafari et al., 2012).

possible to inhibit the connection by phosphorylating sites at the N-terminus of p53, such as S15 (Nakamura et al., 2000). In addition to TP53RK, p53 can be phosphorylated by numerous proteins including checkpoint kinases (Hirao et al., 2000; Shieh et al., 2000), DNA dependent kinases (Lees-Miller et al., 1992), mitogen-activated protein kinases (Milne et al., 1994) and casein kinase 1 (Sakaguchi et al., 2000).

3.1.4.2.2 Serine 15 sites of p53

The serine 15 site of p53 is very important for its regulation (Loughery et al., 2014). This site is phosphorylated by ATM and ATR protein kinases in response to DNA damage (Meek, 2009) and substitution of S15 causes p53 to lose the ability to respond to DNA damage and to activate transcription (Loughery et al., 2014). Activation of transcription by phospho-S15-p53 is thought to be due in part to induction of the histone lysine acetyltransferase activities of p300 and CBP proteins (Dornan & Hupp, 2001). However, the N-terminal domain of p53 is also phosphorylated on other sites, such as Serines 6, 9, 20, 33, 37, 46, and Threonine 18 that are involved to some extent in stabilization, MDM2 binding and transcription (Ashcroft et al., 1999; Sakaguchi et al., 1998).

3.1.4.2.3 Regulation of TP53RK

TP53RK is under the regulation of the Akt/PKB protein kinase (Facchin et al., 2007). Also, TP53RK is regulated and inhibited by CGI-121, a TP53RK binding protein in yeast (Miyoshi et al., 2003).

3.1.4.2.4 KEOPS multiprotein complex

Two independent laboratories have described a group of proteins that are highly conserved and responsible for telomerase maintenance, transcriptional regulation and the modification of tRNA (Srinivasan et al., 2011). One termed it the KEOPS complex (kinase, putative endopeptidase and other proteins of small size) (Downey et al., 2006). Another named It the EKC complex (Endopeptidase-like Kinase Chromatin-associated) (Kisseleva-Romanova et al., 2006), but KEOPS is the most commonly used name in the literature. The KEOPS complex consists of five proteins: LAGE3 (Pcc1p

in yeast), OSGEP (Kae1p in yeast), TP53RK (Bud32p in yeast) and TPRKB (Cgi121p in yeast). A recent study has revealed that reducing the expression of TP53RK slows down the migration of human podocytes, but also accelerates cytoskeletal defects (Braun et al., 2017). Moreover, a gene defect affecting any member of the KEOPS complex results in Galloway–Mowat syndrome, primary microcephaly with nephrotic syndrome (SRNS) (Braun et al., 2017).

3.1.4.2.5 TP53RK and cancer

TP53RK mutants have been reported to occur in cancer as they restrain the pro-apoptotic activity of the protein after mitotic stress. A screen showed that reducing the TP53RK protein may sensitise cancer cells to taxanes (Peterson et al., 2010).

3.1.4.2.6 Properties of the TP53RK homologue in the mouse

The *TP53RK* homologue gene *Trp53rkb* in the mouse has been knocked-out (White et al., 2013) and there was no difference in the percentage or phenotypes of white blood *Trp53rkb*$^{-/-}$ cells at six weeks or at 16 weeks (see Figure 3.1-6 for details). However, the effects of the knockout on sizes of white blood cells was not measured.

3.1.4.2.7 TP53RK protein and cell size

TP53RK is a part of KEOPS complex, as described above and it is highly conserved in multiple species. TP53RK was reported as a cell size regulator in fission yeast (Hayles et al., 2013) and a body size defective protein in *Drosophila* (Björklund et al., 2006). Furthermore, it was reported recently that the protein is required for PI3K/TOR pathway dependent growth in Drosophila (Ibar et al., 2013). This information suggests TP53RK may regulate cell size in more species, such as humans.

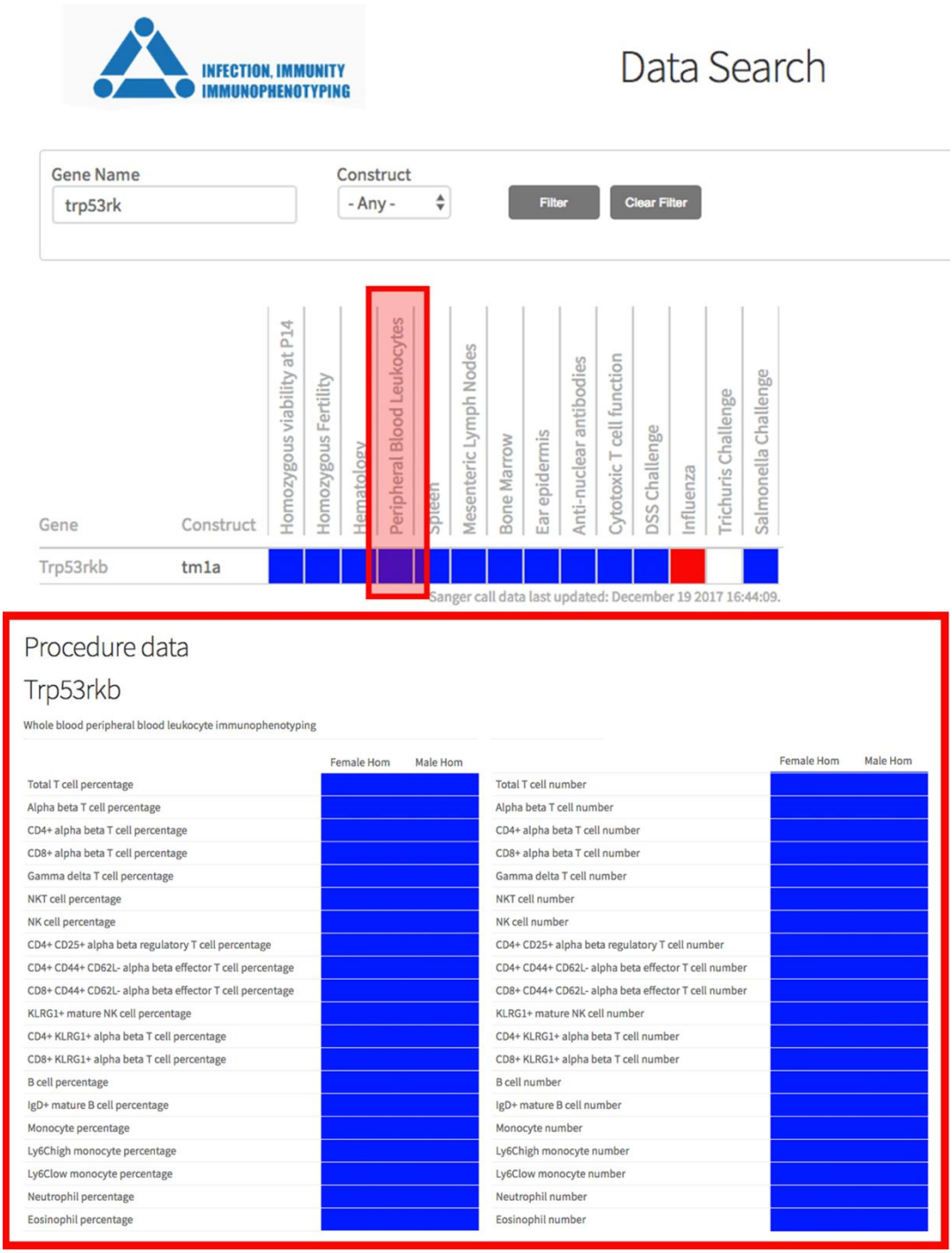


Figure 3.1-6: Phenotype analysis of mouse TP53RK knock out cells

The data shows the effects of mouse **Trp53rkb**$^{-/-}$ as compared with the normal population. The phenotypes of peripheral white blood cells show a normal distribution of percentage and number (colour code: blue = normal; red = different)

Source: http://www.immunophenotype.org/ (White et al., 2013)

3.1.4.3 Prefoldin group in human cells

3.1.4.3.1 Protein folding and chaperones

Protein folding is an active molecular process whereby proteins gain their final multidimensional conformation required for functional activity (Anfinsen, 1973). Molecular chaperonins modulate, assist and stabilise a range of proteins to acquire the necessary conformation (Hartl, 1996). Chaperonins play a role in protein folding, protein assembly, trafficking and degradation that is necessary for development (Hartl et al., 2011). Chaperonins can be categorised into two distinct categories based on their different functions: group I proteins are present in eubacteria and eukaryotic organelles, while group II proteins are present in archaea and eukaryotic cytosol (Yébenes et al., 2011).

Prefoldins are group II chaperonins which fold newly synthesised polypeptide chains, including actins and tubulins (Geissler et al., 1998; Vainberg et al., 1998) and play a fundamental role in the maintenance of cellular homeostasis (Martín-Benito et al., 2007; Siegert et al., 2000). Prefoldins bind proteins as they are translated then transport them to the chaperone Tric/CCT with the help of HSP70 and HSP40 (Geissler et al., 1998; Hartl & Hayer-Hartl, 2002). In eukaryotes, prefoldins form a jellyfish-like heterohexameric protein complex, that consist of six conserved subunits: two α subunits and four β subunits (Siegert et al., 2000). α subunits are PFD3 (commonly named VBP1, gene name *PFDN3*) and PFD5, while β subunits are PFD1, PFD2, PFD4 and PFD6 (Siegert et al., 2000). This unified structure determines the substrate specificity of prefoldins (Abe et al., 2013). Studies have revealed that prefoldins play a crucial role in the cytoskeletal maintenance and cell growth and deletion of individual prefoldin genes slow growth and cause cytoskeletal defects (Geissler et al., 1998; Siegers et al., 1999). Another study suggests that prefoldins play a modifier role by recognising misfolded proteins and reducing cellular toxicity. An example is the recognition of added aggregates of the Huntington protein (Abe et al., 2013; Tashiro et al., 2013).

3.1.4.3.2 Individual members of the prefoldin family

Prefoldin proteins have an individual role in addition to their function as part of a protein complex. *PFDN1* knock-out mice have cerebellar atrophy (Cao et al., 2008). *PFDN5*$^{-/-}$ mice (commonly named as MM-1 α) also have this phenotype in addition to male infertility and the death of Purkinje cells (Lee et al., 2011). The MM-1 α/PFD5 protein is a tumour suppressor and inhibits the function of c-MYC (Mori et al., 1998) that results in the inhibition of cell transformation and cell movement (Fujioka et al., 2001). The PFD2 protein controls the expression of nutrition-related genes by forming a large protein complex with URI and RPB5 (Abe et al., 2013; Gstaiger et al., 2003). VBP1/PFD3 (von Hippel-Lindau-binding protein 1) is the member of the ubiquitin protease system which causes protein degradation (Mousnier et al., 2007; Tsuchiya et al., 1996). VBP1/PFD3 also plays a role in virus replication during HIV-1 infection (Mousnier et al., 2007).

3.1.4.3.3 Prefoldins and the cell size

Some of the prefoldins reportedly affect cell size of budding yeast (Hayles et al., 2013). However, systematic screening has not been carried out to determine whether they also regulate the size of human cells. Therefore, I picked VBP1/PFD3, PFD2, PFD5 and PFD6 to test (see Results section 3.3.4).

3.1.4.4 The MCM7 protein

Eukaryotic DNA replication is rigorously controlled since mistakes cause genomic instability and lead to diseases such as cancer. The pre-replicative complex (Pre-RC) plays a fundamental role in the initiation of DNA replication. Members of this complex include the origin recognition proteins (ORC), CDC6, CDC7, CDT1 and the mini-chromosomal maintenance proteins (MCMs) (Zheng et al., 2017). The latter consists of six homologous proteins that form a hetero-hexamer structure. MCM proteins and the complex are highly conserved in plants as well as in humans (Nieduszynski et al., 2005). Once the MCM helicase proteins form the structure at DNA replication sites, the replication of DNA starts (Blow & Hodgson, 2002), subject to strict control. The mechanisms involved are beyond the scope of this introduction but are reviewed in (Riera et al., 2017).

3.1.4.4.1 Genomic instability in human T-Cells

Work from my laboratory using human peripheral blood T cells reported that reducing MCM levels results in genomic instability, with increased levels of centromere separation, premature chromatid separation and severe chromosome defects including chromosome loss, gain and translocations (Orr et al., 2010). Reducing MCM levels with siRNA in T cells stimulated *via* CD3/CD28 to enter the cell cycle from G_0 results in a reduction in the percentage of cells in S-phase and an increase in cells in G_2 (Orr et al., 2010). The absolute abundance of individual MCM proteins has been quantified and in quiescent (G_0 phase) cells there are low numbers of MCM proteins compared with proliferating cells (Orr et al., 2010; Stoeber et al., 2001). Detecting cells containing high levels of MCM protein expression has been proposed as a means of identifying abnormal proliferating cells in cancers such as cervical and bladder cancers (Dudderidge et al., 2010; Williams et al., 1998).

3.2 Materials and Methods

3.2.1 Testing using human T-cells

The human experiments were carried out exclusively on peripheral blood T-Cells. These T-Cells were isolated from a leukocyte cone of a normal single, healthy, anonymous donor (obtained from the *National Health Service Blood and Transplant Service (NHSBT), Tooting, London, United Kingdom*) with ethical consent.

3.2.2 Location

Experiments with human T-Cells were carried out in the Rayne Institute, Department of Haematological Medicine at the Denmark Hill Campus of King's College London, United Kingdom. All the experimental methods that are explained below are based on published literature and common laboratory usage.

3.2.3 Reagents

All materials were purchased from *Sigma-Aldrich* unless it is otherwise stated. PBS tablets were bought from *Oxiod*, while *methanol* and *ethanol* were bought from *Fisher Scientific*.

3.2.3.1 General reagents

Table 3.2-1. List of general laboratory reagents

Reagent name	Provider
1.5 ml microcentrifuge 'Eppendorf' tubes	StarLabs
ECL-Plus Chemiluminescence detection kit	GE Healthcare
Hybond C-Extra membrane	GE Healthcare
Hyperfilm – ECL	GE Healthcare
Novex Sharp Pre-Stained Protein Standard (Western Blot Marker)	InVitrogen
NuPAGE 4-12% (w/v) polyacrylamide BisTris gels	InVitrogen
NuPAGE LDS sample buffer (4x)	InVitrogen
NuPAGE Running Buffer	InVitrogen
NuPAGE Transfer Buffer	InVitrogen
Saran Wrap	Commercially available from Sainsbury's
Skimmed milk powder	Commercially available, Marvel
Trizol	InVitrogen
Whatman 3MM paper	VWR International Ltd

3.2.3.2 Tissue culture reagents and equipment

Table 3.2-2. List of tissue culture reagents and equipment

Reagent name	Provider
15 ml and 50 ml (Falcon) centrifuge tubes	VWR International Ltd
15 ml Pipettes	VWR International Ltd
6 well plates (TPT plates were used for siRNA transfections)	VWR International Ltd
Anti CD3/CD28 immunomagnetic beads	InVitrogen
Foetal calf serum (FCS)	InVitrogen
Hematopoietic Media (X-VIVO)	InVitrogen
Histoplaque 1077	InVitrogen
Human serum	Invitrogen
Human T-Cell Nucleofection Kit (including Transfection Buffer)	Lonza
siRNA	Dharmacon or SIGMA (see Table 3.2-9)
T Cell Negative Isolation Kit	InVitrogen
Vented tissue culture flasks	VWR International Ltd

3.2.3.3 Enzymes, Solutions and Buffers

Table 3.2-3. List of enzymes, solutions and buffers

Reagent name	Provider
SDS lysis buffer (4x)	125 mM Tris-HCl pH 6.8, 4% (w/v) SDS, 40% (v/v) Glycerol, 200 mM DTT
Cell cycle stain	40µg/ml Propidium Iodide, 5µg/ml Fluorescein isothiocyanate and 1µg/ml of RNase A in PBS
MES SDS Running Buffer for Western Blots	50mM MES, 50mM Tris, 0.1% (w/v) SDS, 1mM EDTA, pH 7.3
MES SDS Transfer Buffer for Western Blots	25mM Bicine, 25mM Bis-Tris, 1mM EDTA, pH 7.2 plus 20% (v/v) Methanol
Phosphate Buffered Saline (PBS)	1 tablet (phosphate buffer, potassium chloride 0.02% (w/v), sodium chloride 0.8%(w/v)) dissolved per 100ml of H_2O
Primary antibody incubation buffer	3% (w/v) BSA, PBS, 0.05% (v/v) Tween-20
RNase A	10 mg/ml in dH_20
Secondary antibody incubation buffer	10% (w/v) Non-fat dried milk, PBS, 0.05% (v/v) Tween-20
Western Blot blocking solution (PBST)	10% (w/v) Non-fat dried milk, PBS, 0.05% (v/v) Tween-20
Western Blot wash buffer	PBS, 0.05% (v/v) Tween-20

3.2.3.4 Equipment

Table 3.2-4. List of major equipment

Name	Provider
Azure C300 chemiluminescent imager	Azure Biosystems
Becton Dickinson FACS Caliber Canto II - Flow cytometer	BD Biosciences
Cellometer Auto T4 Cell Counter	Nexcelom Bioscience
FlowJo Software (version 9)	Flow Jo Inc
Light microscope with lens 4x-10x-40x-100x	Nikon
Nucleofector 2b Device	Lonza

3.2.4 Isolation of Quiescent T-Cells

Quiescent T-Cells were isolated from a single-donor leukocyte cone from a healthy, human donor. The cone was delivered in the morning to the laboratory; then the following protocol was followed.

The white cells in the leukocyte cone were diluted 1:1 with Phosphate Buffered Saline (PBS) (Dulbecco) and gently pipetted into 50 ml tubes. These diluted cells were layered over an equal volume of *Histopaque 1077* (Sigma) in 50ml (Falcon) centrifuge tubes, then centrifuged for 30 minutes at 560 × G_{max} with no brake. Thereafter, PBMCs that form a layer at the Histopaque/PBS interface was removed with a Pasteur pipette and resuspended in 50 ml PBS, then centrifuged at 400 × G_{max} for 10 minutes. The remaining platelets were removed by resuspending the cell pellet in 50 ml PBS with the

addition of 2% (v/v) Foetal Bovine Serum (FBS) (Sigma) and centrifugation at 200 × G_{max}. This process was carried out three times.

After the third wash, the numbers of live and dead cells were determined by manual counting. A small aliquot of PBMC were mixed with 0.4% (w/v) Trypan blue solution (Sigma) (1:1 (v/v) ratio) and visualised using a light microscope (Nikon) and live/dead cell numbers were determined by counting non-stained/blue cells using a Neubauer Improved counting chamber (0.0025 mm^2 resolution). A full leukocyte cone yielded around 1 billion (1×10^9) PBMC. For my experiments, I usually used 2.5-5 × 10^8 PBMC from which to isolate T-Cells.

Non-activated, quiescent T-Cells were isolated from the PBMC by negative selection using the T-Cell Negative Isolation Kit (Invitrogen). The following antibodies are present in the kit: anti-CD14, anti-CD16 (a and b), anti-CD56 and CD235a. I also supplemented the kit with anti-HLA class II (DR and DP) to remove any activated T-Cells. The antibodies were incubated with the PBMCs at 4°C for 20 mins with constant rolling. The T-Cells were collected by centrifugation at 400 × G_{max} for 8 mins and then washed twice with room temperature PBS, 2% (v/v) FCS and centrifugation at 400 × G_{max} for 8 mins. The pelleted cells were then resuspended in 50 ml PBS, 2% (v/v) FCS and were incubated with anti-immunoglobulin-conjugated magnetic beads (Invitrogen) for 20 minutes at room temperature with constant rolling. Vigorous pipetting with a 5 ml pipette dispersed bead clumps. Cells other than non-activated T-Cells bind to the beads and were removed using a magnetic tube holder (Invitrogen). In the supernatant, the unbound cells are a mixture of $CD4^+$ and $CD8^+$ quiescent T-Cells. Bead removal was repeated if it was necessary after microscopic inspection. The non-activated, quiescent T-Cells were then seeded in X-VIVO media (Invitrogen) at 1-4 × 10^6 cells/ml, then kept in a tissue-culture incubator in a fully humidified 5% CO_2 atmosphere at 37°C until further use. Experiments with isolated T-Cells were started within 24 hours of the isolation.

Each full T-Cell isolation kit typically yielded 1.25 - 2.5 × 10^8 non-activated T cells. For those individual experiments that I have performed. I typically used ¼ or ½ of a full negative isolation kit. After that, typically I used between 75 – 150 million (1 × 10^6) T-Cells for my experiments.

3.2.5 Stimulation of quiescent (G_0) T-Cells

In my experiments, T-Cell stimulation was needed, and the quiescent T-Cells were stimulated by the addition of Dynabeads Human T-Activator CD3/CD28 magnetic beads (ThermoFisher Scientific) at 0.5 bead/cell (10 µl beads/1 × 10^6 cells).

Before using the magnetic beads, they were washed with sterile PBS three times, twice with 20 µl PBS, then with 10 µl PBS per 10 µl beads. Beads were isolated from each wash step by using a magnetic tube holder, according to the manufacturer's protocol (Invitrogen).

3.2.6 Quantifying the percentage of cells in each cell cycle phase

Flow cytometric analysis was used to analyse the cell cycle profile of a population of T-Cells. 2 × 10^5 cell samples were taken at the time points indicated in the Results section. The cells were petted by centrifugation at 200 × G_{max}., then fixed in 400 µl of 70% (v/v) ethanol at -20°C. Then the samples were transferred to Fluorescence Activated Cell Sorter (FACS) tubes, and the fixed cells were centrifuged at 400 × G_{max} for 8 minutes. The supernatant was discarded. The cell pellet was resuspended in 400 µl Fluorescein Isothiocyanate (FITC) / Propidium Iodine (PI) cell cycle stain (40 µg /ml PI, 5 µg /ml FITC, 1 µg / ml RNase 1, sterile PBS) and incubated for 30 minutes at 37°C. Subsequently, flow cytometric analysis was performed using a Becton Dickinson FACS Caliber Canto II machine. At least 10,000 events were recorded by the machine for each target population (raw version of **Figure 3.2-1 part C**).

To analyse cell cycle data, I used **Flow Jo software** (version 10.0.7r2, https://www.flowjo.com/) to display two-dimensional dot plots and density plots of flow cytometric data with different scatter and emission properties.

By plotting **FSC-A** (Forward Scatter - Area) against **SSC-A** (Side Scatter - Area) the healthy T-Cell population was selected (lymphocyte electronic "gate") (see **Figure 3.2-1 part A**). Then, by plotting **Compensated-PE-A** (Phycoerythrin – Area, detected the emission of red florescent Propidium Iodine, Ex_{max} 495 nm / Em_{max} 565 nm) against **PE-W** (Side Scatter - Width) doublets were eliminated from the analysis (see **Figure 3.2-1 part B**). Single cells were selected in this population of **Compensated-FITC** (Ex_{max} 495 nm / Em_{max} 520 nm) plotted against **Compensated-PI** (see **Figure 3.2-1 part C**) and electronic "gates" were applied to quantify the numbers of cells in each cell cycle phase according to DNA (PI-stained) and protein (FITC-stained) content: G_0, G_1, S-phase, G_2/M and Sub G_0/Apoptotic cells.

This method to discriminate single, healthy T-Cells in the sample is based on the method published earlier by Thomas laboratory (Lea, Buggins, et al., 2003).

The distribution of T-Cells in the gates described above was calculated then CSV (comma separated value) files of single cells (selected population in **Figure 3.2-1 part B**), and all cell cycle phases (**Figure 3.2-1 part C)** were exported for further data analysis.

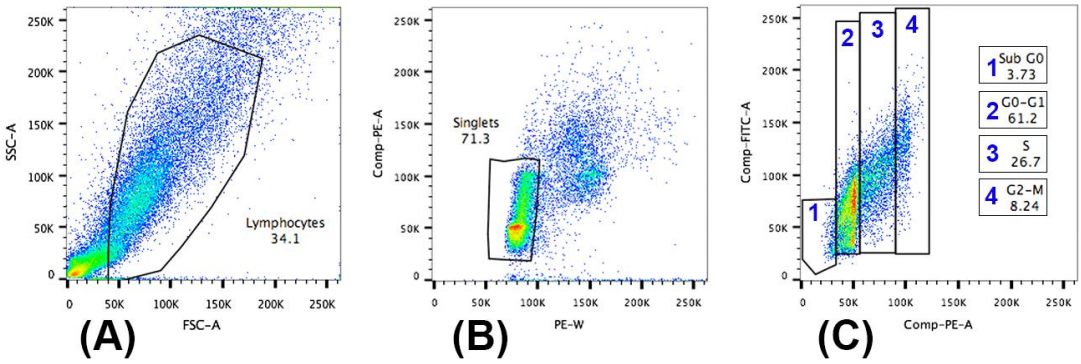


Figure 3.2-1: Density plots of CD3/CD28 stimulated T-Cells.

T-Cells were stimulated, then a sample was taken 72h after CD3/CD28 stimulation. The cells were fixed, stained with PI/FITC and analysed by flow cytometry, as described in the text. Each dot on the plots depicts a two-parameter display of data for a single cell (the parameters for each plot are shown on the X- and Y-axes, as detailed below). The data of cells in the area (electronic "gate") in panel (A) are analysed in (B), and the cells in the electronic gate in (B) are analysed in (C).

Part A: SSC-A was plotted against FSC-A to select healthy T-Cells.

Part B: This is a sub-population of Part A. Compensated PE-A was plotted against PE-W to select and analyse single cells. Cell doublets were eliminated from further analysis as they are outside the electronic gate.

Part C: This is a sub-population of Part B. Gates for individual cell cycle phases are shown, and the number of cells in each gate is calculated

In this given example, the distribution of cells is as follows:

(1) Sub G_0/Apoptotic cells 3.73%,

(2) G_0 - G_1 combined phases: 61.2 %,

(3) S-phase: 26.7 % and

(4) G_2/M phase: 8.24 %.

3.2.6.1 Analysis of Flow Cytometric Data

CSV files containing FACS data, such as FSC, SSC, FITC, PE, Time and Event number were exported and analysed using FlowJo software. After that, a script written in PHP programming language using PHPExcel extension (https://github.com/PHPOffice/PHPExcel) was used to individually clear, trim and concatenate different experiments and time points to two single XLSX excel file sheets. Table 3.2-5 indicates which data source was used in the Excel files.

Two different export files were made:

- one for each experiment and time points comparing control experiment side by side,
- one for each time point comparing all experiments that were performed in parallel from the same donor.

Table 3.2-5: List of elements of FlowJo CSV export files

Name	Area	Height	Width
Forward Scatter (FSC)	**Used**	not used	not used
Side Scatter (SSC)	**Used**	not used	not used
Fluorescein Isothiocyanate (FITC)	not used	not used	not used
Compensated - FITC	**Used**	n/a	n/a
Phycoerythrin (PE)	not used	not used	not used
Compensated - PE	**Used**	n/a	n/a

3.2.6.1.1 Processing exported CSV files with Microsoft Excel

XLSX files were processed in Microsoft® Excel for Mac (version 15) software. The raw data was separated from the sheets analysed; then a statistical calculation was made to display the basic statistics of data:

- Number of events incorporated in each phase,
- Minimum value,
- Maximum value,
- Average (Arithmetic mean),
- Standard deviation
- Standard error and
- Coefficient of Variation values.

Distribution of all events at a given time point was calculated, and histograms are presented as bar and line charts. I used 4 fluorescent light channel values as described in Table 3.2-6 binned for histogram analysis. These values and results were compared to the control population using *f*-test and Student's *t*-test.

Table 3.2-6: List of bins and ranges for histograms

Name	Bin_{min}	Bin_{max}	Bin size
FSC A	**42,000**	180,000	2,000
SSC-A	**10,000**	240,000	5,000
Compensated - FITC A	**10,000**	240,000	5,000
Compensated - PE A	**32,000**	110,000	2,000

3.2.6.1.2 Flow Chart of the analysis method

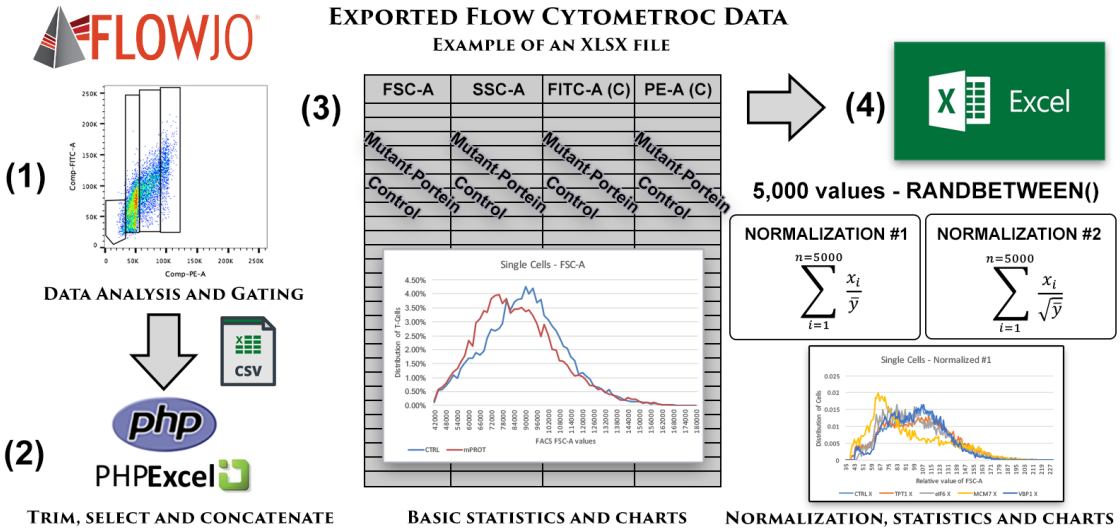


Figure 3.2-2: Flow Chart of the methods used to analyse the Flow Cytometric Data

(1) FACS data were first analysed, gated in FlowJo software, then exported to individual CSV files for single cells and cell cycle phases: G_0, G_1, S-phase, G_2/M and Sub G_0/Apoptotic cells.

(2) Data CSV files were trimmed, cleared and concatenated with PHP code using the phpExcel extension, then exported to XLSX files.

(3) Basic statistics and charts were created for FSC-A, SSC-A, Compensated FITC-A and Compensated PE-A.

(4) Then 5,000 events were randomly selected with the RANDBETWEEN() function; two randomisation methods were introduced. Finally, line and bar charts were created.

3.2.6.1.3 Normalisation and randomisation

To compare individual experiments with a different number of events and from different donors, normalisation and randomisation were introduced. Using Microsoft® Excel for Mac (version 15) software, 5,000 events were selected at random (with RANDBETWEEN() function) for each set of data being compared (including at least one control siRNA experiment), while two normalisation methods were used:

(1) Each set of data points were divided by the mean of the control population of the parallel experiment.

$$\sum_{i=1}^{n} \frac{x_i}{\bar{y}}$$

- where **n** = 5,000, the number of randomly selected events
- **x**, individual values of FACS data for experimental data
- **y**, as individual values of FACS data for control(s)

(2) The second normalisation method Involves division by square root means of the related control population.

$$\sum_{i=1}^{n} \frac{x_i}{\sqrt{\bar{y}}}$$

- where **n** = 5,000, the number of randomly selected events
- **x**, individual values of FACS data for experimental data
- **y**, individual values of FACS data for control data

3.2.6.2 CD3 purity test

CD3 is only expressed on the surface of T-cells and not on other cell types in the blood, and so the percentage of cells that express CD3 in any given population can be used to determine the percentage of T-Cells present. Therefore, the CD3 analysis was used to determine the purity of T-Cells isolated by negative selection. The method used a CD3 APC-H7 antibody (BD Bioscience) and was carried out within 1h of isolation. 2×10^4 cells were removed from the isolated T-Cell population into each of two round-bottomed FACS tubes, then each cell sample was washed with 50 ml sterile PBS and centrifuged at 200 × G_{max} for 5 minutes. The supernatant was poured off (approximately 100 µl PBS is left in each tube after pouring off the bulk of the supernatant), and 5 µl of CD3 APC-H7 antibody was added to one of the two tubes. The samples were incubated for 15 minutes in the fridge (temperature 5° C). T-Cell samples were centrifuged again at 200 × G_{max} for 5 minutes, then 200 µl sterile PBS was added and flow cytometric analysis was performed using a Becton Dickinson FACS Caliber Canto II machine. The emission of the CD3 APC-H7, APC-H7-A (Ex_{max} 650 nm/Em_{max}785 nm) antibody was plotted against SSC-A to determinate the purity of T-Cells present in the sample (+CD3 APC-H7 *versus* no antibody control).

3.2.7 Culturing T-Cells

The non-activated, quiescent (G_0) T-Cells isolated by negative selection were cultured at 1 - 4 × 10^6 cells/ml in X-VIVO media with 2% (v/v) FBS. Cells proliferate in response to CD3/CD28 stimulation and once proliferating cell numbers double every 23h (Lea, Orr, et al., 2003). The proliferating cells were kept in the range 5×10^5 to 1×10^6 by the addition of cell culture medium, as necessary.

3.2.8 Protein lysates and western blotting

1×10^6 cells were pelleted at 200 × G_{max} for 5 mins and then lysed in SDS buffer solution or NuPAGE LDS sample buffer solution (ThermoFisher Scientific - Invitrogen) (see above; 50 µl per 1×10^6 cells). 10 µl of each protein sample was loaded onto each lane

of a 4-12% (w/v) polyacrylamide Bis-Tris gel (ThermoFisher Scientific - Invitrogen), and 10 µl of Novex Protein Standards (ThermoFisher Scientific - Invitrogen) was also run in an appropriate lane.

Proteins were separated by gel electrophoresis at 200 V / 120 mA for 1 h (1 × running buffer: 50 mM MES, 50 mM Tris, 1 mM EDTA, 0.1% (w/v) SDS, pH 7.3). The proteins separated by electrophoresis in the gel were transferred by electroblotting at 25 V / 160 mA for 1.5 h onto a nitrocellulose membrane (Hybond C Extra, GE Healthcare) (1 × transfer buffer: 25mM Tris-HCl, 25 mM bicine, 1 mM EDTA, pH 7.3, 20% (v/v) methanol). The remaining binding sites on the nitrocellulose membrane were blocked by incubation in PBS, 10% (w/v) non-fat dried milk (skimmed milk powder, Marvel) for 30 mins at room temperature. Before antibody probing, the blot was washed quickly three times with 100-200 ml PBS, 0.05% (v/v) Tween-20.

3.2.9 Antibody probing

Primary antibodies were typically used at dilutions of 1:1000 in PBS, 3% (w/v) BSA, 0.01% (w/v) NaN_3. The final concentrations of primary antibodies were between 0.1 - 1 µg/ml. The blots were incubated in primary antibodies overnight, using 50 ml tubes and 5 ml primary antibody solution. Subsequently, the membranes were washed 3 × 5 mins in 100ml PBS, 0.05% (v/v) Tween-20 and incubated in the appropriate secondary antibody for 30 mins. Then the blots were washed three times in 100 ml PBS, 0.05% (v/v) Tween-20 and protein bands were detected using the ECL-Plus detection system (GE Healthcare). Blots were wrapped in Saran Wrap and exposed to X-ray film (Hyperfilm–ECL, GE Healthcare) for various periods of time and developed using a Compact X4 X-Ray Developer (Konica-Minolta) (until January 2016). After that, chemiluminescent images were acquired using a digital scanner (Azure C300 chemiluminescent digital imager).

I used the following primary antibodies stated in Table 3.2-7.

3.2.9.1 Primary Antibodies used for Western Blotting

Table 3.2-7: List of primary antibodies used for Western Blotting

Primary	Clone or Identifier	Dilution	Species	Manufacturer
C14orf166	ab49342	1:1000	Rabbit	Abcam
eIF6	#3263	1:1000	Rabbit	New England Biolabs
GAPDH	ab8245	1:1000	Mouse	Abcam
Histone H3	#4499	1:500	Rabbit	Cell Signalling
MCM4	AHP840	1:1000	Goat	Bio-Rad
MCM7	DCS-141	1:1000	Mouse	Sigma
p107 (C-18)	sc-318	1:500	Rabbit	Santa-Cruz Biotechnology
p130	sc-317	1:500	Rabbit	Santa-Cruz Biotechnology
p53	sc-6243	1:1000	Rabbit	Santa-Cruz Biotechnology
PFDN2	ab206691	1:1000	Rabbit	Abcam
PFDN2	ab156146	1:1000	Rabbit	Abcam
PFDN5	sc-19843	1:1000	Goat	Santa-Cruz Biotechnology
PFDN5	ab129116	1:1000	Rabbit	Abcam

PFDN6	sc-376733	1:1000	Mouse	Santa-Cruz Biotechnology
pRB	ab39689	1:1000	Rabbit	Abcam
phospho-RB (S^{805}/S^{811})	#9308S	1:500	Rabbit	Cell Signalling
phospho-p53 (S^{15})	#9286	1:1000	Mouse	Cell Signalling
TNPO3	ab109386	1:1000	Rabbit	Abcam
TP53RK	sc-85846	1:1000	Goat	Santa-Cruz Biotechnology
TPT1 / TCTP	#5128	1:1000	Rabbit	Cell Signalling
VBP1	#19837	1:1000	Mouse	Cell Signalling
VPS18	ab178689	1:1000	Rabbit	Abcam

3.2.9.2 HRP-Conjugated Secondary Antibodies used for Western Blotting

Table 3.2-8: List of secondary antibodies used for Western Blotting

Secondary	Clone or Identifier	Stock concentration	Source Species	Manufacturer
Anti-Mouse	P0447	1:2000	Goat Polyclonal	Dako
Anti-Rabbit	P0448	1:2000	Goat Polyclonal	Dako
Anti-Goat	P0449	1:1000	Rabbit Polyclonal	Dako

3.2.10 T-Cell transfections with siRNA

Quiescent, non-stimulated T-Cells were transfected within 24 hours of isolation. The selected amount of quiescent, non-stimulated T-Cells was centrifuged at 200 × G_{max} for 5 minutes, and the supernatant was discarded. To transfect T-Cells a *Nucleofector*® device and Nucleofection solutions (Lonza) were used. The siRNA used for each target was either a pool of four siRNA (Dharmacon, GE-Healthcare) or custom-made single siRNA (Sigma), based on published sequences. 5 nmol of the Dharmacon siRNA was diluted with 50 µl RNase-free water to 500 µmolar concentration. For the custom-made siRNA, 10 nmol was diluted with 100 µl RNase-free water to 500 µmolar concentration. From both, 5 µl siRNA per 5 million (1×10^6) T-Cells was added to the cell pellet with 100 µl Nucleofection reagent (Human T-Cell Nucleofector Kit, Lonza, cat # VPA-1002). Then, the T-Cell/siRNA mix was transfected to a cuvette and Nucleofection was carried out using an *Amaxa Nucleofector 2B device* (Lonza) using programme U-014. Transfected cells were collected gently with a micropipette then gently transferred to pre-warmed X-VIVO media and left for 24-72 hours to recover at 37° C in the tissue-culture incubator. The viable T-Cells were then counted manually using a *Neubauer Improved* cell counting chamber (0.0025 mm^2 resolution) in the presence of 0.4% (w/v) Trypan blue solution (Sigma)) (1:1 (v/v) ratio) and visualised using a light microscope (Nikon). Typically, ~35-40% of the original population survived the Nucleofection procedure. To stimulate the transfected cells, Dynabeads Human T-Activator CD3/CD28 magnetic beads were added to the culture media at 0.5 bead/cell (10 µl beads/1×10^6 cells).

3.2.10.1 siRNA sequences

Table 3.2-9: List of siRNA sequences

These sequences were ordered from Dharmacon and used as stated in the table:

(1) used as SMARTpool mix or

(2) used as SMARTpool mix and individual sequences

(for TPT1 only)

The SMARTpool mix contains all four of the related targeting sequences.

Catalogue Name	Product ID	Used Product	Protein Name	Target Sequence
siGENOME Non-Targeting siRNA Pool #1	D-001206-13	SMARTpool	Control	UAGCGACUAAACACAUCAA
siGENOME Non-Targeting siRNA Pool #1	D-001206-13	SMARTpool	Control	UAAGGCUAUGAAGAGAUAC
siGENOME Non-Targeting siRNA Pool #1	D-001206-13	SMARTpool	Control	AUGUAUUGGCCUGUAUUAG
siGENOME Non-Targeting siRNA Pool #1	D-001206-13	SMARTpool	Control	AUGAACGUGAAUUGCUCAA
siGENOME human EIF6 (3692) siRNA	D-010096-01	SMARTpool	eIF6	GAGCUUCGUUCGAGAACAA
siGENOME human EIF6 (3692) siRNA	D-010096-02	SMARTpool	eIF6	GAUCGGAGGCUCAGAGAAC
siGENOME human EIF6 (3692) siRNA	D-010096-04	SMARTpool	eIF6	CGAGAACAACUGUGAGAUC
siGENOME human EIF6 (3692) siRNA	D-010096-05	SMARTpool	eIF6	CAAUUGAAGACCAGGAUGA
siGENOME human MCM4 (4173) siRNA	D-003275-05	SMARTpool	MCM4	GGACAUAUCUAUUCUUACU
siGENOME human MCM4 (4173) siRNA	D-003275-06	SMARTpool	MCM4	GAUGUUAGUUCACCACUGA

siGENOME human MCM4 (4173) siRNA	D-003275-07	SMARTpool	MCM4	CCAGCUGCCUCAUACUUUA
siGENOME human MCM4 (4173) siRNA	D-003275-08	SMARTpool	MCM4	GAAAGUACAAGAUCGGUAU
siGENOME human MCM7 (4176) siRNA	D-003278-05	SMARTpool	MCM7	GGAAAUAUCCCUCGUAGUA
siGENOME human MCM7 (4176) siRNA	D-003278-08	SMARTpool	MCM7	GGAGAGAACACAAGGAUUG
siGENOME human MCM7 (4176) siRNA	D-003278-21	SMARTpool	MCM7	UGUCAUACAUUGAUCGACU
siGENOME human MCM7 (4176) siRNA	D-003278-22	SMARTpool	MCM7	GAGACAAUGACCUACGGUU
siGENOME human PFDN2 (5202) siRNA	D-013011-01	SMARTpool	PFDN2	GAACAACAAGGAGCAGAUA
siGENOME human PFDN2 (5202) siRNA	D-013011-02	SMARTpool	PFDN2	GGAGCGAACUGUCAAAGAG
siGENOME human PFDN2 (5202) siRNA	D-013011-03	SMARTpool	PFDN2	CACAGCAGCUUCAGGCAAA
siGENOME human PFDN2 (5202) siRNA	D-013011-04	SMARTpool	PFDN2	GGUGAUUGCUGGCUUCAAC
siGENOME human PFDN5 (5204) siRNA	D-011352-01	SMARTpool	PFDN5	UAACCAAGCAGAUGGAGAA
siGENOME human PFDN5 (5204) siRNA	D-011352-02	SMARTpool	PFDN5	UCAAGAGGAAGAUAGAUUU
siGENOME human PFDN5 (5204) siRNA	D-011352-03	SMARTpool	PFDN5	GAACUGGGUACUAUGUAGA
siGENOME human PFDN5 (5204) siRNA	D-011352-04	SMARTpool	PFDN5	UAGAAAUGCUCAAGAACCA
siGENOME human TPT1 (7178) siRNA	D-004559-01	SMARTpool, Individual	TPT1	AGAUGUUCUCCGACAUCUA
siGENOME human TPT1 (7178) siRNA	D-004559-05	SMARTpool, Individual	TPT1	GGUUGUGCCUGGAGGUGGA

siGENOME human TPT1 (7178) siRNA	D-004559-06	SMARTpool, Individual	TPT1	UCUACAAGAUCCGGGAGAU
siGENOME human TPT1 (7178) siRNA	D-004559-16	SMARTpool, Individual	TPT1	CGAAGGUACCGAAAGCACA
siGENOME human VBP1 (7411) siRNA	D-011419-01	SMARTpool	VBP1	GGAGGAAGACCUUGACUUU
siGENOME human VBP1 (7411) siRNA	D-011419-02	SMARTpool	VBP1	GAACAGUACCAGAAGUAUA
siGENOME human VBP1 (7411) siRNA	D-011419-03	SMARTpool	VBP1	GAACUCAACCUUGCUCAAA
siGENOME human VBP1 (7411) siRNA	D-011419-04	SMARTpool	VBP1	GGGCUAAUGUAAUGCUUGA

3.2.10.2 Custom-made siRNA sequences

Table 3.2-10: List of custom-made siRNA sequences for PFDN2 and PFDN5.

These sequences were published and used in the literature by two different publications by a research group from Hokkaido University, Japan (Abe et al., 2013; Miyazawa et al., 2011)

Protein Name	Producer	Sense Sequence	Antisense Sequence
PFDN2	Sigma	GUCUGAACGUGCUGAACAAGA	UUGUUCAGCACGUUCAGACAG
PFDN5	Sigma	GCCUAGUGAUCGAUACACUGA	AGUGUAUCGAUCACUAGGCUG

3.2.11 Analysing and counting T-Cells with a Cellometer

The analysis was carried out to analyse the viability and measure the cell size profile of a population of T-Cells pre and post CD3/CD28 stimulation, as described in the Results section.

1×10^5 T-Cell samples were taken, then centrifuged at $200 \times G_{max}$ for 5 minutes. Subsequently, the supernatant was discarded, and the cell pellet was resuspended in 10 µl sterile PBS. Samples were mixed with 0.4% (w/v) Trypan blue solution (Sigma) (1:1 (v/v) ratio), thereafter 10 µl was pipetted into SD100 Slides (Nexelcom Biosciences). Slides were analysed with a Cellometer Auto T4 Cell Counter (Nexelcom Bioscience), then data was exported to CSV files for further analysis.

3.2.12 Analysis of cell size profiles by light microscopy

To analyse cell sizes of T-Cell populations, samples of 5×10^4 cells were taken, then centrifuged at $200 \times G_{max}$ for 5 minutes. Subsequently, the supernatant was discarded, and the cell pellet was gently resuspended in 20 µl sterile PBS. The samples were immobilised onto poly-lysine microscopic slides using a Shandon Cytospin 4 (ThermoFisher Scientific) machine. After the cytospin had finished, the slides were taken to King's College Hospital NHS Trust Foundation, Department of Haematology to stain them with May-Grünwald Giemsa (MGG) dye. This dye stain cells so that the sub-cellular structures are visible by light microscopy. Digital light microscope (Nikon) images were acquired, and cell sizes were analysed using NIS-Elements software.

3.3 Results

3.3.1 T-Cells: entry into the Cell Cycle from the quiescent state

Peripheral blood T-Cells remain in a quiescent state in the circulation until they encounter a mitogenic stimulus (Thomas, 2004). Quiescent T-Cells are small and have 2n DNA content (*Ibid*). In response to CD3/CD28 stimulation (Surh & Sprent, 2008) they activate effector functions, enter the growth cycle and become bigger and they enter the cell cycle and divide (Lea, Orr, et al., 2003).

In my experiments I used the DNA replication protein MCM7 as a positive control for cell cycle regulation. This protein is not present or present at low levels in quiescent T cells and is induced in response to CD3/CD28 stimulation (Lea, Orr, et al., 2003). I also used the ribosome biogenesis protein, eIF6 as a positive control. It is also induced in response to CD3/CD28 and reducing the induction of eIF6 by transfecting quiescent T cells with siRNA prior to CD3/CD28 stimulation affects cell size (Orr et al., 2012).

3.3.1.1 Stimulation of quiescent T-Cells

3.3.1.1.1 CD3 purity test

Primary, quiescent, non-activated T-Cells were isolated from peripheral blood PBMCs by negative selection. The purity was checked by determining the percentage of cells expressing CD3 on their surface (CD3 APC H7 antibody, Figure 3.3-1). Panel B shows cells stained with the CD3 APC H7 antibody and panel A shows the no-antibody control (used to set the negative electronic "gate"). The Figure (Panel B) shows an analysis of a typical T-Cell isolate in which approximately 96.4% of the cells express CD3 (98.8%-2.4%). Since the negative gate (no antibody control, Panel A) is conservatively set at approximately 2%, I conclude that the sample contains pure T-Cells. This type of analysis was carried out on every T cell isolate used in my study.

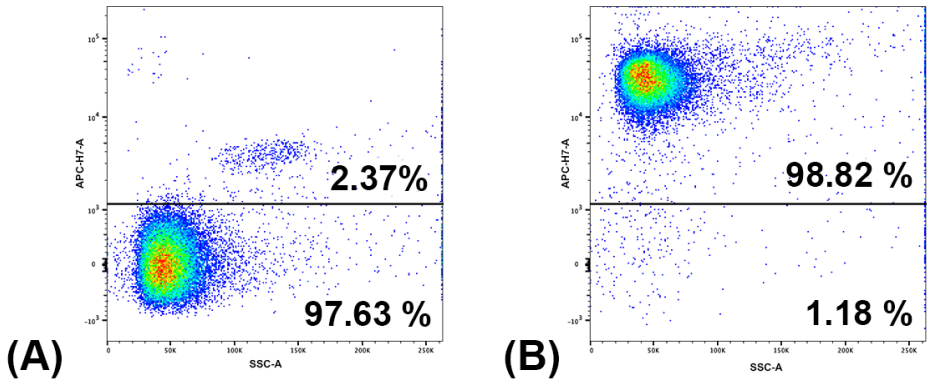


Figure 3.3-1: CD3 purity test of isolated quiescent T-Cells

Quiescent, non-stimulated T-Cells were isolated from a healthy human donor. Analysis to check T-Cell purity was carried out by analysing the percentage of cells expressing CD3 on their surface:

(A) T-Cells incubated **without** CD3 APC-H7 antibody. This is the negative control.

(B) T-Cells incubated **with** CD3 APC-H7 antibody.

Samples were analysed by flow cytometry, and the percentage of cells in each sector (electronic "gate") is shown.

3.3.1.1.2 Cell cycle analyses by flow cytometry

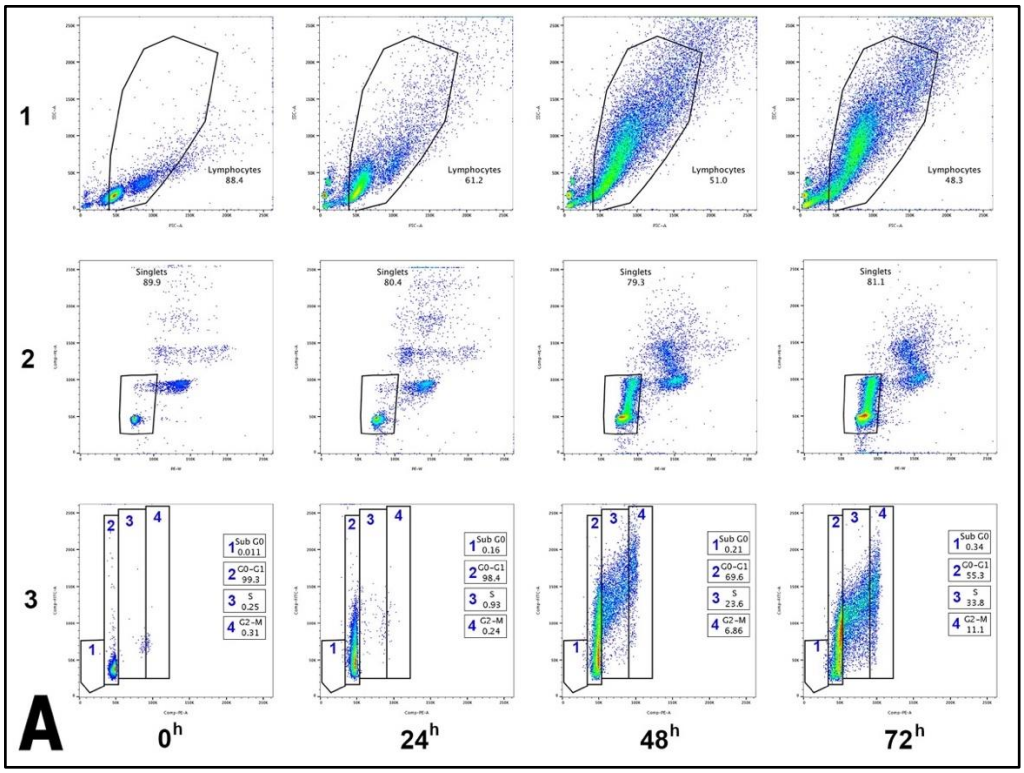
T-Cells isolated as described above were stimulated with CD3/CD28 beads. Samples were taken at different time points after adding the beads to the cell culture media: 0, 24, 48 and 72h post stimulation. Cell samples were fixed and stained at each time point with PI (for total DNA content) and FITC (for protein content), then analysed by Flow Cytometry. The method quantifies the DNA and protein content of each cell, as described in (Lea, Orr, et al., 2003). The data are shown in Figure 3.3-2 A and the percentage of cells in each cell cycle phase are quantified in B. In parallel, Western-Blot samples were taken to check the phosphorylation of the retinoblastoma protein, which becomes phosphorylated on S807/811 in response to CD3/CD28 stimulation (Lea, Orr, et al., 2003) (Figure 3.3-2 C).

In Figure 3.3-2 A there are three rows of flow cytometry data arranged into four columns for samples collected at each of the four different time points (0, 24, 48 and 72h post stimulation). In the first row, flow cytometry data is shown, as FSC-A (Forward Scatter - Area) plotted against SSC-A (Side Scatter - Area). Debris in the media was excluded from analyses by setting an electronic gate that includes only the lymphocyte population. Then single cells were selected by gating Compensated-PE-A (Phycoerythrin – Area) against PE-W (Side Scatter - Width), shown in the second row. Compensated-FITC (protein content on the y-axis) was plotted against Compensated-PI (DNA content on the x-axis) for this population, shown in the third row. Electronic gates were set to quantify the percentage of cells in each cell cycle phase: G_0, G_1, S-phase, G_2/M and Sub G_0/Apoptotic cells. The percentages for this experiment are shown in B. The data show that initially, 99.30% of cells are in a quiescent state and fewer than 1% in S or G_2/M phases. At 24h post stimulation, cells have not started to replicate DNA (0.93% reached S phase and 0.24% G_2/M phases). T-Cells do not respond synchronously to CD3/CD28 stimulation and the leading front of CD3/CD28 stimulated cells enter S-phase approximately 35h post stimulation and 23.60% are in S phase and 6.86% G_2/M phases 48h post stimulation. By 72h, 33.80% are in S phase and 11.10% G_2/M phases 72h. Note that there is a difference in the protein content of T cells stimulated with CD3/CD28 for 24h, as compared with the 0h pre-stimulation

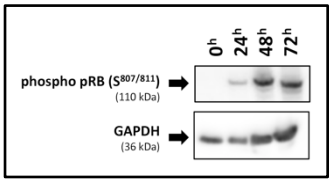
sample of quiescent (G_0) cells. Previous work by the laboratory (Lea, Orr, et al., 2003) showed that T-Cells become committed to entering the cell cycle and growth cycle 2-5h post stimulation and at 24h post-stimulation the leading front of cells are in late G_1 and are larger than those in G_0. The kinetics of entry into the cell cycle and the growth cycle from quiescence in response to CD3/CD28 shown here agrees with work published previously by our laboratory (Lea, Orr, et al., 2003).

The retinoblastoma protein (pRb) is hypo-phosphorylated in quiescent cells and becomes phosphorylated on many different sites, including Serine 807 and 811 in response to CD3/CD28 stimulation (Lea, Orr, et al., 2003; Thomas et al., 1998). Phosphorylation at $S^{807/811}$ in T-Cells is detectable by western blotting by 16h post-stimulation. The phosphorylation state of pRb in the T-Cell samples described above was analysed by Western blotting and the data (Figure 3.3-2 C) show that phosphorylation of $S^{807/811}$ is not detectable in unstimulated T-Cells and becomes detectable in the samples stimulated for 24h and longer with CD3/CD28, consistent with previously published data.

The main objective of the project is to analyse the effects of specific proteins on cell size. To do that, I tested several methods for measuring the sizes of T-Cells.



B



C

Time	0h	24h	48h	72h
G_0-G_1	99.30%	98.40%	69.60%	55.30%
S	0.25%	0.93%	23.60%	33.80%
G_2-M	0.31%	0.24%	6.86%	11.10%

Figure 3.3-2: Cell cycle analyses and western blot analyses of CD3/CD28-stimulated T-Cells

(A) T-Cells were stimulated with CD3/CD28 beads. Samples were taken at times shown after CD3/CD28 stimulation: 0, 24, 48 and 72h. The samples were fixed, stained with PI (DNA content) and FITC (protein content) and analysed by flow cytometry. In all-time points shown single, T-Lymphocytes were analysed by setting the appropriate electronic gates.

(B) The percentage of cells in each cell cycle phase for each sample is shown.

(C) Western blot of phospho-pRb($S^{807/811}$) and GAPDH[7] (used as the loading control).

[7] **Glyceraldehyde 3-phosphate dehydrogenase**, an enzyme that is constitutively expressed and considered as a housekeeping protein.

3.3.1.1.3 Cell size measurements by Flow Cytometry

The flow cytometer measures the way the laser light reflects from each cell as it passes through the beam. The forward scatter (FSC-A) measurement correlates well with cell size (Miettinen & Björklund, 2016). A previous study by my laboratory showed that this measurement increases as T-Cells increase in size as they progress through the G_1 phase of the cell cycle in response to CD3/CD28 stimulation (Orr et al., 2012). Therefore, I used the FSC-A data from my flow cytometry analyses shown in Figure 3.3-2 to determine how this measurement changes with time after CD3/CD28 stimulation. To do this, I used FSC-A data extracted for cells in all phases of the cell cycle and the histogram of the distribution shows that pre-stimulation (0h) the quiescent cells have a small, tight FSC-A distribution and this increases with time after CD3/CD28 stimulation up to 48h post stimulation. Some T-Cells have divided 72h post stimulation (Lea, Buggins, et al., 2003) and the size profile is smaller for this sample (Figure 3.3-3). The mean and range of FSC-A for each sample are shown in Table 1.3-1 and discussed In Section 3.3.1.2.1 below. Exact size data can be obtained from the FSC-A measurement if a reference standard, such as 10 μm polystyrene microparticles are analysed in the same experiment (Yurinskaya et al., 2017). However, this reference was published too late for most of my analyses and I did not used a reference sample for FSC-A in my study.

Further to analysing the FSC-A intensity values, I checked the protein content of cells stained with FITC (FITC-A intensity values) (Orr et al., 2012). In Figure 3.3-4, I show the distribution of protein content of T-Cells pre and post stimulation *via* CD3/CD28. The FITC values increase when the T-Cells are stimulated with CD3/CD28 beads in line with the increase in FSC-A values, in agreement with previously published work by my laboratory (*ibid*).

In further siRNA experiments described in this Chapter used this method to analyse the sizes of T-Cells in different cell cycle phases as they progress from quiescence into the first cell cycle. In my cultures, almost all the T-Cells analysed have not divided (Lea, Orr, et al., 2003) and see Section 3.2.6.2 However, the method of analysing FSC-A or

FITC content is limited for analysing actively dividing T-Cells as it does not distinguish between cells that have divided one, twice or more times. This could be important as T-Cells have a limited capacity to divide and once they reach the Hayflick number of cell divisions (Hayflick, 1965) they enter replicative senescence and cannot divide further (Perillo et al., 1989). There may be cell size differences between actively dividing cells having undergone different numbers of cell divisions and those entering replicative senescence. To overcome this, it might be possible to develop a method that also incorporates a dye such as CFSE or other cell permeable dyes which my laboratory has used to analyse the division of T-Cells(Kordasti et al., 2016; Lea, Buggins, et al., 2003). However, because I only analysed T-Cells that are predominantly in the first cell cycle it was not necessary to develop such a method.

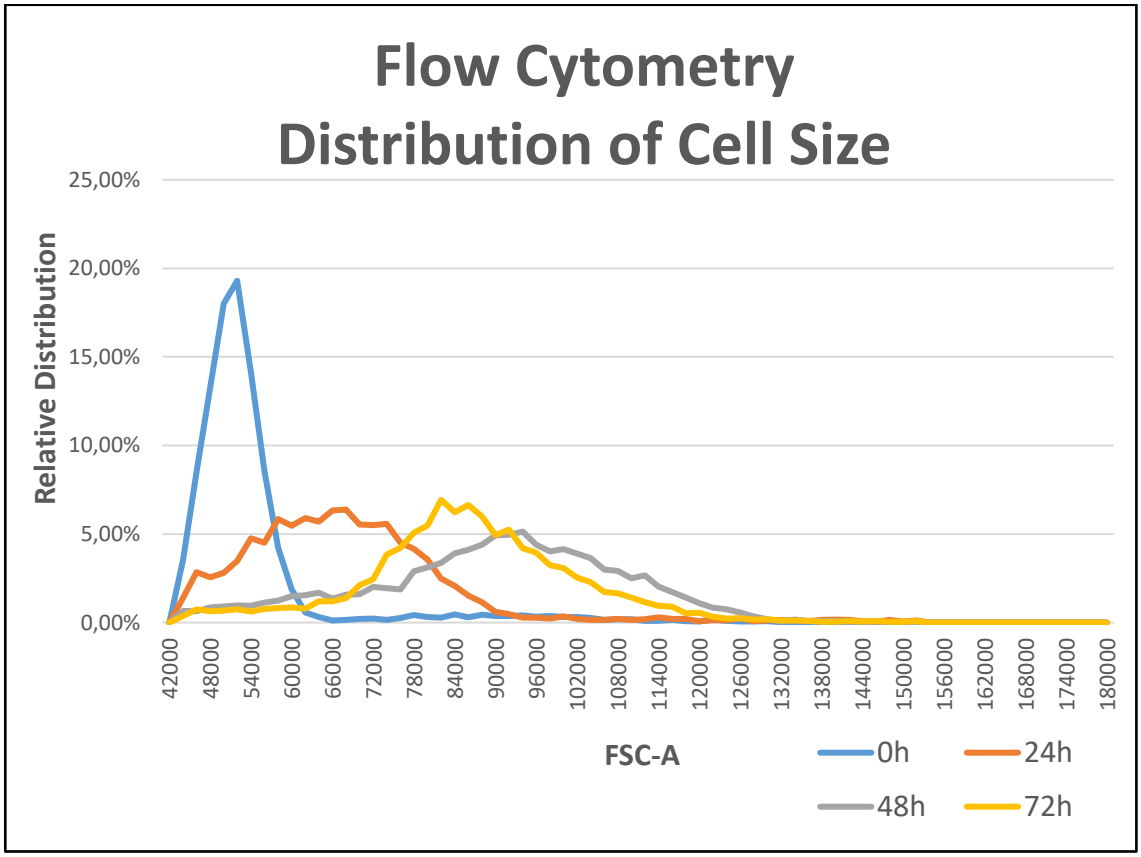


Figure 3.3-3: Histogram chart of the distribution of cell sizes, based on flow cytometry data

FSC-A intensity values are plotted on the x-axis, while the y-axis shows the relative distribution of T-Cells at the time-points shown. These data are derived from analyses shown in Figure 3.3-2 A.

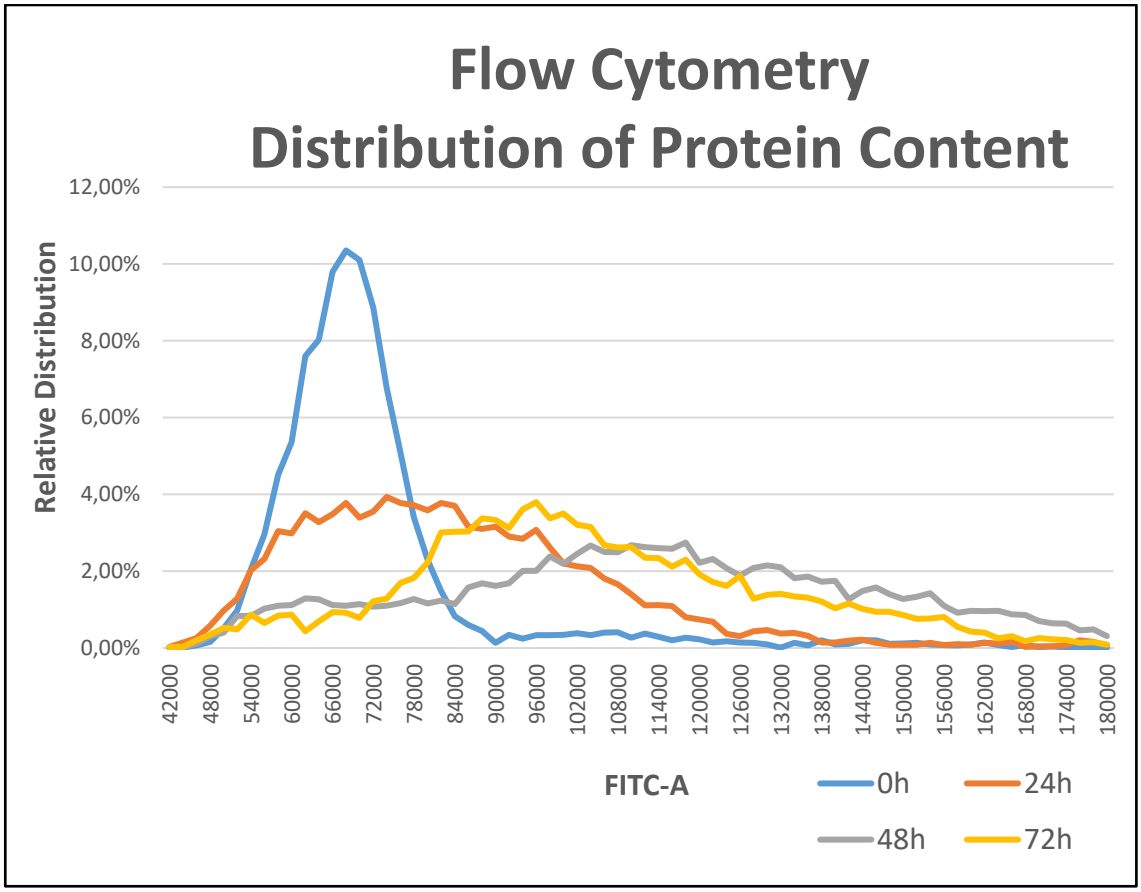


Figure 3.3-4: Histogram chart of the distribution of protein content, based on flow cytometry data

FITC-A intensity values are plotted on the x-axis, while the y-axis shows the relative distribution of T-Cells at the time-points shown. These data are derived from the same analyses shown for FITC-A intensity values in Figure 3.3-3 and Table 3.3-1.

3.3.1.1.4 Cell size measurements by other methods

In addition to FSC-A measurements, I also tested two other methods for measuring cell size: automated Cellometer and manual light microscopy analyses of MGG-stained cells.

3.3.1.1.4.1 Cellometer analyses

Samples of T-Cells were cultured as described for Figure 3.3-3, stained with Trypan blue and analysed using an automated Cellometer. The Cellometer is an automated image analyser which automatically recognises, counts and measures the diameter of cells. Images of the cells analysed by this method are shown in Figure 3.3-5 and the cell measurements are quantified below. Note that quiescent T-Cells are single cells, but CD3/CD28 stimulation results in the expression of adhesion molecules that are important for immune responses and which cause activated T-Cells to form clumps (Bierer & Burakoff, 1988; Dustin, 2001). By 24h post CD3/CD28 stimulation and subsequently the Cellometer images obtained show that T-Cells have formed clumps. The T-Cells which form clumps are the cells that have been stimulated with CD3/CD28 and therefore those most important for my analyses. This clumping therefore complicates analyses of cell size and the Cellometer was not able to distinguish individual cells, and I could not adequately disaggregate the clumps into single cells by manual pipetting with a p1000 Gilson pipette (the method used routinely in the laboratory). It might be possible to disaggregate clumps of cells by using a protease such as Trypsin. However, there is a risk of causing cell death and such a method has not been developed in the laboratory.

3.3.1.1.4.2 Light Microscopy analyses

Samples of T-Cells (as described for Figure 3.3-3) were taken at the time points shown and immobilised on microscope slides by Cytocentrifugation. The immobilised cells were fixed and stained with May-Grünwald Giemsa (MGG) stain, which enables individual cells to be visualised and analysed under light microscopy (Figure 3.3-6). Cell numbers and measurements of cell sizes were carried out of these images using ImageJ software, and these data are described below.

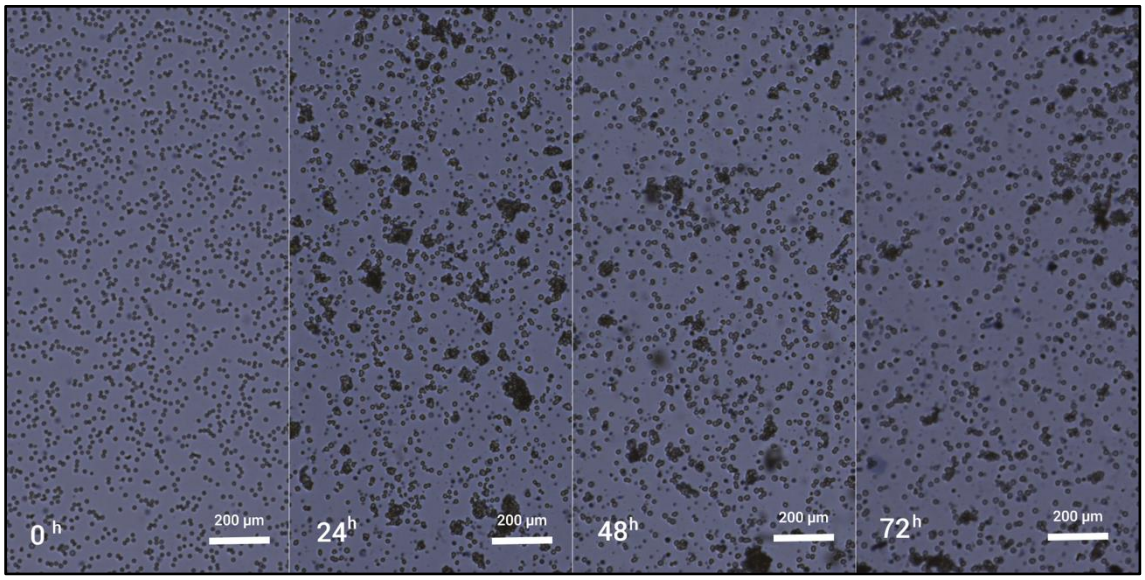


Figure 3.3-5: Cellometer images of CD3/CD28-stimulated T-Cells

T-Cells were stimulated as described for Figure 3.3-2 and analysed using an automated Cellometer, using standard settings.

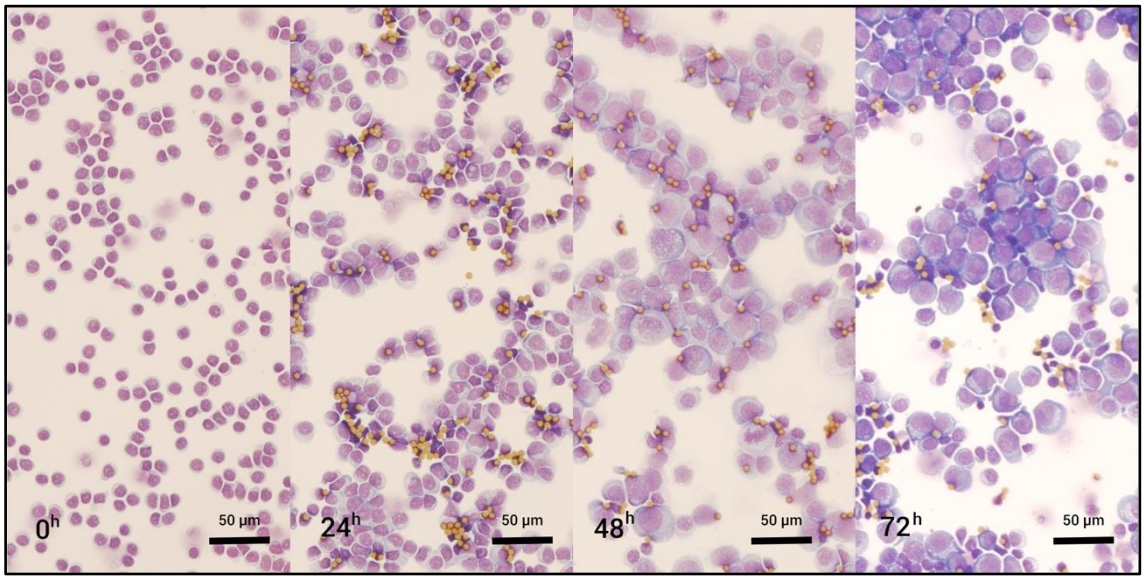


Figure 3.3-6: Light microscopy images of MGG-stained T-Cells

T-Cells were stimulated as described for Figure 3.3-2; samples were taken at times shown after stimulation, immobilised on microscope slides and stained with MGG stain. The cells were visualised under light microscopy, and cell numbers and measurements of cell size were performed manually with ImageJ software. Small orange spheres are the CD3/CD28 beads used to stimulate the T-Cells.

3.3.1.2 Comparing different methods for analysing cell size

As described above, three different approaches were used to count and measure the size of non-stimulated and CD3/CD28-stimulated T-Cells. The data obtained in each case was used to compare and select the best method for further use.

3.3.1.2.1 Flow Cytometry Data – Forward Scatter Area

I acquired FSC-A data, as described above and basic statistics of the flow cytometry. These data show that quiescent cells (0h) have a small, tight FSC-A distribution and the mean and range of FSC-A values increase after CD3/CD28 stimulation. Cells double in size during each cell cycle so that the size of cells in the population are maintained after each cell division. As described in the Introduction Section 3.1.2.4.1 quiescent T-Cells increase in size but not DNA content by 24h post CD3/CD28 stimulation as they enter G_1 and start to enter S-phase by 30-48h. Cell cycle entry is not synchronous, and this results in a heterogeneity of cell sizes post stimulation, as discussed in (Orr et al., 2012). By 24h, <1% of T-Cells have divided, which increases to 16.4% by 48h. 36% have divided once, and 1% have divided twice at 72h (Lea, Orr, et al., 2003). Work published previously by our laboratory also showed that FSC-A increases in line with the increase in protein content of T-Cells as they enter G_1 (Orr et al., 2012) and my data are consistent with the increase in cell size that occurs as T-Cell enter and progress through the cell cycle.

3.3.1.2.2 Cellometer data

Automated Cellometer images were analysed using standard settings and the data are shown in Table 3.3-2. The average sizes of cells at individual time points did not change with time post CD3/CD28 stimulation and remained at approximately 12 μm. Also, the standard deviation and coefficient of variation values were very high, and the cell numbers counted in the same amount of each sample varied from 1.6 to 4 thousand. T-Cells significantly increase in size as they progress through the cell cycle. The Cellometer data did not reflect this change. As it can be seen on Figure 3.3-5, it is likely that the analyses are simply of single cells that have not yet responded to CD3/CD28 stimulation since the stimulated, activated T-Cells form large clumps, as described above that I could not adequately disaggregate. Therefore, the Cellometer was not used in further studies in my Thesis.

3.3.1.2.3 Analysis data of MGG staining

Images were analysed using the oval selection tool of ImageJ software, which was used to analyse pictures taken from MGG stained slides. Standard settings were used, and the data are shown in Table 3.3-3. I obtained very similar results to those using the Cellometer: the average sizes of cells at individual time points did not change with time post CD3/CD28 stimulation and remained at approximately 13 μm. In addition, the standard deviation and coefficient of variation values were very high.

The cell numbers counted in the same amount of each sample varied from 120 to 150, which is low as compared with the 10,000 cells analysed by flow cytometry. This is a manual counting method with software quantification, but there is high variability because of the differences between fields of the cell samples. T-Cells increase in size as they progress into and through the cell cycle, which is apparent when different fields of cells are viewed, but the manual measurement of cell sizes on the MGG slides did not reflect this phenomenon. Since it would be very time-consuming to quantify multiple fields of cells to obtain representative data of sufficient numbers of cells, I did not use this method in further analyses.

Table 3.3-1: Basic statistics of flow cytometry FSC-A data

Time	Number of Events	Minimum	Maximum	Mean	Standard Deviation	Coefficient of Variation	Signal to Noise Ratio
G0	7,504.00	42,498.90	140,255.09	53,597.45	12,507.87	23.34%	4.29
24h	6,944.00	42,498.90	151,918.20	67,267.85	15,946.23	23.71%	4.22
48h	7,555.00	42,505.20	151,651.80	88,872.34	18,687.35	21.03%	4.76
72h	8,067.00	42,597.00	151,824.59	85,391.81	15,812.08	18.52%	5.40
Average	7,517.50	42,525.00	148,912.42	73,782.36	15,738.38	21.65%	4.66

Table 3.3-2: Basic statistics of cell size, based on automated Cellometer data

Time	Number of Events	Minimum	Maximum	Mean	Standard Deviation	Coefficient of Variation	Signal to Noise Ratio
G0	1,882.00	4.74 µm	39.65 µm	11.32 µm	5.09 µm	107.29%	0.93
24h	1,675.00	4.11 µm	38.13 µm	12.74 µm	5.25 µm	127.67%	0.78
48h	3,432.00	4.11 µm	39.01 µm	12.00 µm	4.53 µm	110.18%	0.91
72h	3,963.00	4.11 µm	39.83 µm	11.97 µm	4.73 µm	115.01%	0.87
Average	2,738.00	4.27 µm	39.16 µm	12.01 µm	4.90 µm	115.04%	0.87

Table 3.3-3: Basic statistics of cell size, based on the analysis of MGG stained slides

Time	*Number of Events*	*Minimum*	*Maximum*	*Mean*	*Standard Deviation*	*Coefficient of Variation*	*Signal to Noise Ratio*
G_0	149.00	6.22 µm	35.76 µm	12.99 µm	6.38 µm	102.58%	0.97
24h	153.00	4.27 µm	33.56 µm	13.76 µm	5.92 µm	138.75%	0.72
48h	141.00	4.74 µm	37.65 µm	13.17 µm	5.65 µm	119.22%	0.84
72h	120.00	5.43 µm	38.33 µm	13.39 µm	5.87 µm	108.13%	0.92
Average	140.75	5.17 µm	36.33 µm	13.33 µm	5.96 µm	117.17%	0.86

3.3.2 Expression of the individual proteins predicted by Bioinformatics analyses in quiescent and CD3/CD28-stimulated T cells.

In Chapter 2, I used Bioinformatics analyses of other species to predict proteins that may regulate the size of human cells. To be able to carry out experiments using human T-Cells, the proteins must be expressed and be detectable in this cell type. The expression of each of the proteins was analysed in samples of quiescent as well as CD3/CD28 stimulated T-Cells by western blotting (Figure 3.3-7). I selected the proteins based on the bioinformatic analysis I described earlier, and the selection criteria are discussed in Chapter 2, section 2.3.2 .

The figure shows western blotting data of three different individual T-Cell isolates and in each case the kinetics of cell cycle entry was checked (one experiment is shown in **panel A**).

In **panel B**, I analysed the expression of the TP53RK protein kinase, which phosphorylates p53 on serine 15. The data show that TP53RK is not expressed in quiescent T-Cells but is induced in CD3/CD28-stimulated cells. p53 is also induced by CD3/CD28 (unpublished data from our laboratory) and the western blot shows that p53 is induced and phosphorylated on S^{15}. The expression of the DNA replication protein MCM7 was analysed as a control since it is induced during cell cycle entry (Orr et al., 2010).

I analysed the expression levels of the prefoldin proteins, PFD2, VBP1/PFD3 and PFD5 and TCTP (Figure 3.3-7 C and D). Moreover, I checked the expression of a known cell size regulator, eIF6 as published previously by my laboratory (Orr et al., 2012), as well as Mcm7. All proteins are expressed in stimulated T-Cells, while they are only present in quiescent cells at low or undetectable levels. The data show that TP53RK, Prefoldin proteins and TCTP are all expressed in human T-Cells after stimulation with CD3/CD28 beads and are clearly detectable by western blotting.

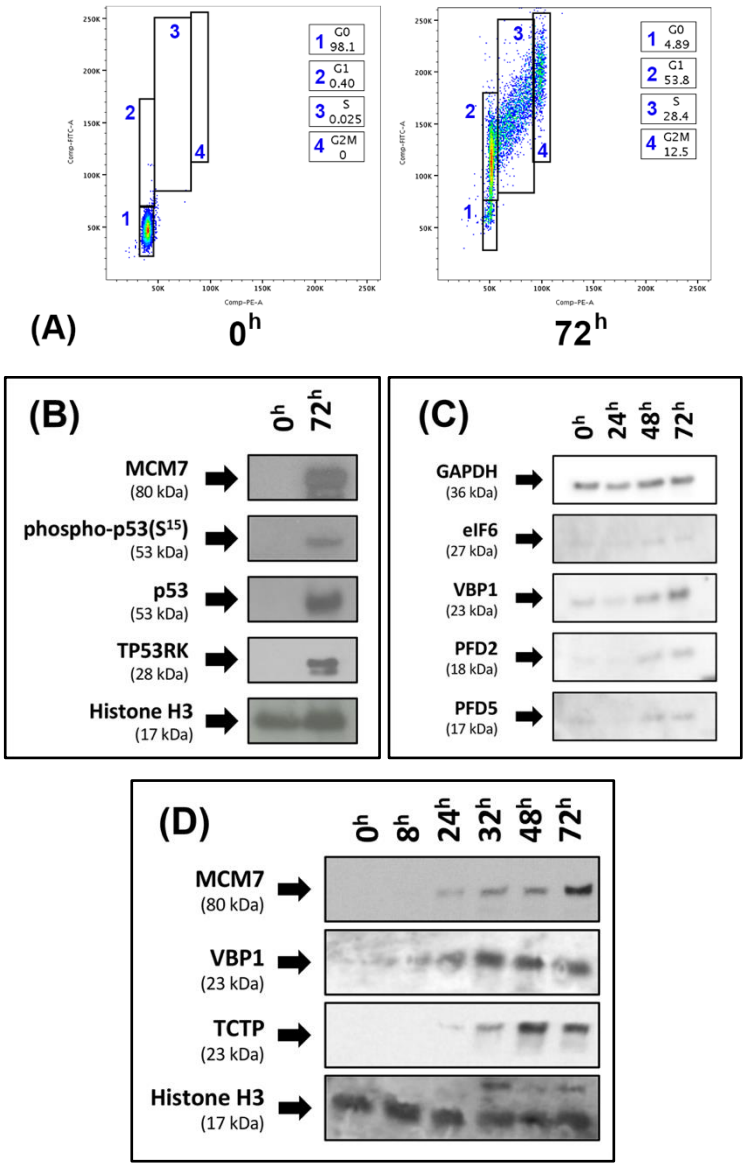


Figure 3.3-7: Expression of individual proteins in quiescent and CD3/CD28 stimulated T-Cells

Quiescent T-Cells were Isolated and stimulated with CD3/CD28 for the time-points shown. Samples were prepared for cell cycle analyses and western blotting as described in Sections 3.2.8 **(A)** Cell cycle analysis: cells were fixed and stained with PI and FITC, and the percentage of cells in each cell cycle phase was determined by flow cytometry. A representative of the n=3 T cell isolates used for experiments shown in this figure. **(B)** Western blot of TP53RK. The blot was also probed for p53, phospho-p53(S^{15}) and Mcm7. **(C)** Western blot of VBP1 PFD2, PFD5, eIF6 and **(D)** TPT1, VBP1 and MCM7. For each blot, sample loading was checked either by analysing the expression of Histone H3 or GAPDH.

3.3.3 Reducing the induction levels of individual proteins in T cells with siRNA.

The experiments in the last section show that the proteins predicted as potential cell size regulators by my Bioinformatics analysis, namely TP53, the prefoldin proteins and TCTP, are all expressed at low or undetectable levels in quiescent T-Cells and are induced when the T-Cells are stimulated *via* CD3/CD28. In previous experiments in our laboratory, the effects of reducing the levels of specific proteins were determined by transfecting quiescent T-Cells with siRNA and then stimulating with CD3/CD28 (Orr et al., 2010, 2012). The presence of siRNA in the cells prevented the induction of the specific protein that would normally occur in response to CD3/CD28 stimulation. I intended using the same method in my experiments on cell size by determining whether preventing the induction of the protein of interest changes the percentage of cells in individual cell cycle phases and/or affects cell size.

3.3.3.1 Reducing expression of MCM complex proteins

Before Investigating the effects of reducing the expression of the proteins predicted by my Bioinformatic analyses, I repeated some of the siRNA experiments that have been published previously by my laboratory, namely reducing the expression of the MCM proteins Mcm4 and Mcm7 (Orr et al., 2010).

I isolated non-activated peripheral blood T-Cells, checked their quiescence as described in section 3.2.6.2 , then transfected these cells with MCM4 and MCM7 siGENOME SMARTpool siRNAs by Nucleofection (as described in section 3.2.10). All T-Cells in the population take up the siRNA (Lea, Buggins, et al., 2003). The transfected, quiescent T cells were cultured without stimulation for 3 days to allow them to recover and for low levels of Mcm7 (and Mcm4) in the quiescent cells to reduce, then stimulated with CD3/CD28 beads for 72 h. Both Mcm4 and Mcm7 siRNA transfections were carried out with duplicate or triplicate isolates of T-Cells. Samples were collected for western blotting, to determine whether the siRNA transfection had reduced Mcm4 and Mcm7 expression and for cell cycle analyses, to determine the percentage of cells in each cell cycle phase. Western blot results (Figure 3.3-8 panel B)

show that transfection with MCM4 siRNA reduces the expression of Mcm4 protein compared with the control siRNA. Similarly, transfection with MCM7 siRNA reduces Mcm7 protein expression. Differences in Mcm4 (and Mcm7) expression between samples of control-siRNA and MCM4-siRNA transfected cells were determined by scanning the images and quantification with ImageJ software. Transfection of MCM4 siRNA reduced Mcm4 protein expression by approximately 50% and MCM7 siRNA reduced Mcm7 protein expression by 85% or more, compared with T-Cells transfected with non-targeting control siRNA. These data show that the T-Cells were transfected with the siRNA and that the siRNA reduced the CD3/CD28-stimulated induction of Mcm4 or Mcm7, as published previously (Orr et al., 2010).

To determine the effects on the cell cycle of transfecting T-Cells with MCM4 or MCM7 siRNA, the percentage of cells in each cell cycle phase was determined by flow cytometry (Figure 3.3-8 panel A). The data show that reducing Mcm4 or Mcm7 reduces the percentage of cells in S-phase, the same percentage are in G_2/M and more remain in G_0/G_1. These data are consistent with (Orr et al., 2010). The cells in G_2/M are probably due to cell cycle arrest in G_2 as cells exit S-phase with DNA damage (*ibid*), although this was not investigated in my study.

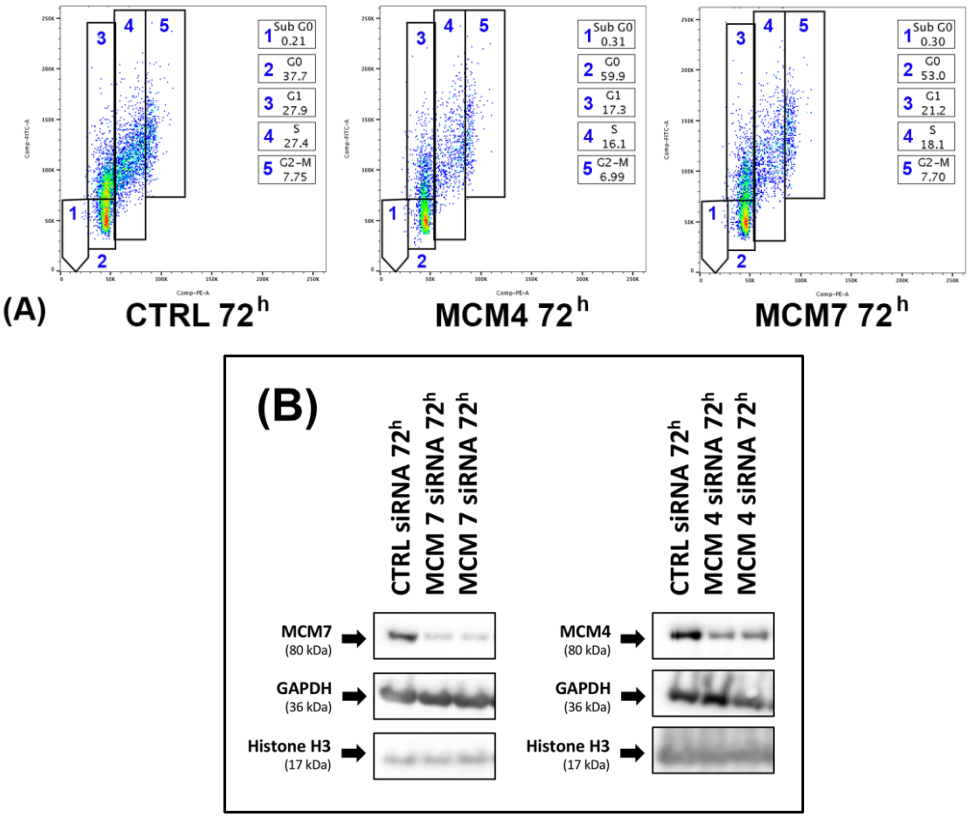


Figure 3.3-8: Transfection of T-Cells with MCM4 and MCM7 siRNA

Non-activated, quiescent T-Cells were transfected with MCM4 siRNA, MCM7 siRNA or non-targeting control siRNA and stimulated with CD3/CD28 for 72h. Samples were taken for cell cycle analyses by flow cytometry and for western blotting, to determine the expression of Mcm4 and Mcm7 proteins.

(A) Samples were fixed, stained with PI and FITC and the percentage of cells in each cell cycle phase was determined by flow cytometry.

(B) Western blots of T-Cells transfected with the siRNAs indicated were probed for the expression of Mcm4 or Mcm7 proteins. For both siRNA experiments, blots were also probed for GAPDH and Histone H3 as a loading control.

3.3.3.2 Reducing the expression of proteins of interest

3.3.3.2.1 Western Blotting

In the section above, I showed that I could reduce the expression of two proteins in T-Cells, namely Mcm4 and Mcm7 by transfecting siRNA and that this resulted in changes in cell cycle progression. Therefore, I tested whether I could use the same siRNA approach to reduce expression levels of proteins that were predicted by my bioinformatic analysis (Chapter 2.3.2). I carried out experiments with SMARTpool siRNA for TCTP, TP53RK and VBP1. Western blots of these experiments are shown in Figure 3.3-9 panel A. The TCTP and VBP1 siRNA reduced the expressions of Tctp and Vbp1 proteins respectively to very low or undetectable levels. The reduction in Tctp was to a consistently low level, but the reduction of Vbp1 was variable. In contrast, TP53RK siRNA did not reduce Tp53rk protein expression in five independent experiments. Experiments were also carried out with twice the amount of TP53RK siRNA, but Tp53rk levels were not reduced and therefore, I dropped this protein from further analyses. I also carried out experiments transfecting T-Cells with PFD2 and PFD5 siRNA SmartPools, but again the levels of these proteins were not reduced. I also tested custom siRNAs based on sequences published by a research group from Hokkaido University (Abe et al., 2013) and I also used a different antibody (see Section 3.2.9.1) to detect the expression levels of both proteins. However, neither the SmartPools nor the published siRNA caused a reduction of Pfd2 or Pfd5 protein expression (Figure 3.3-9 panel C). In these experiments, siRNA to MCM7 and EIF6 (reported by (Orr et al., 2012)) both caused a reduction in the corresponding proteins.

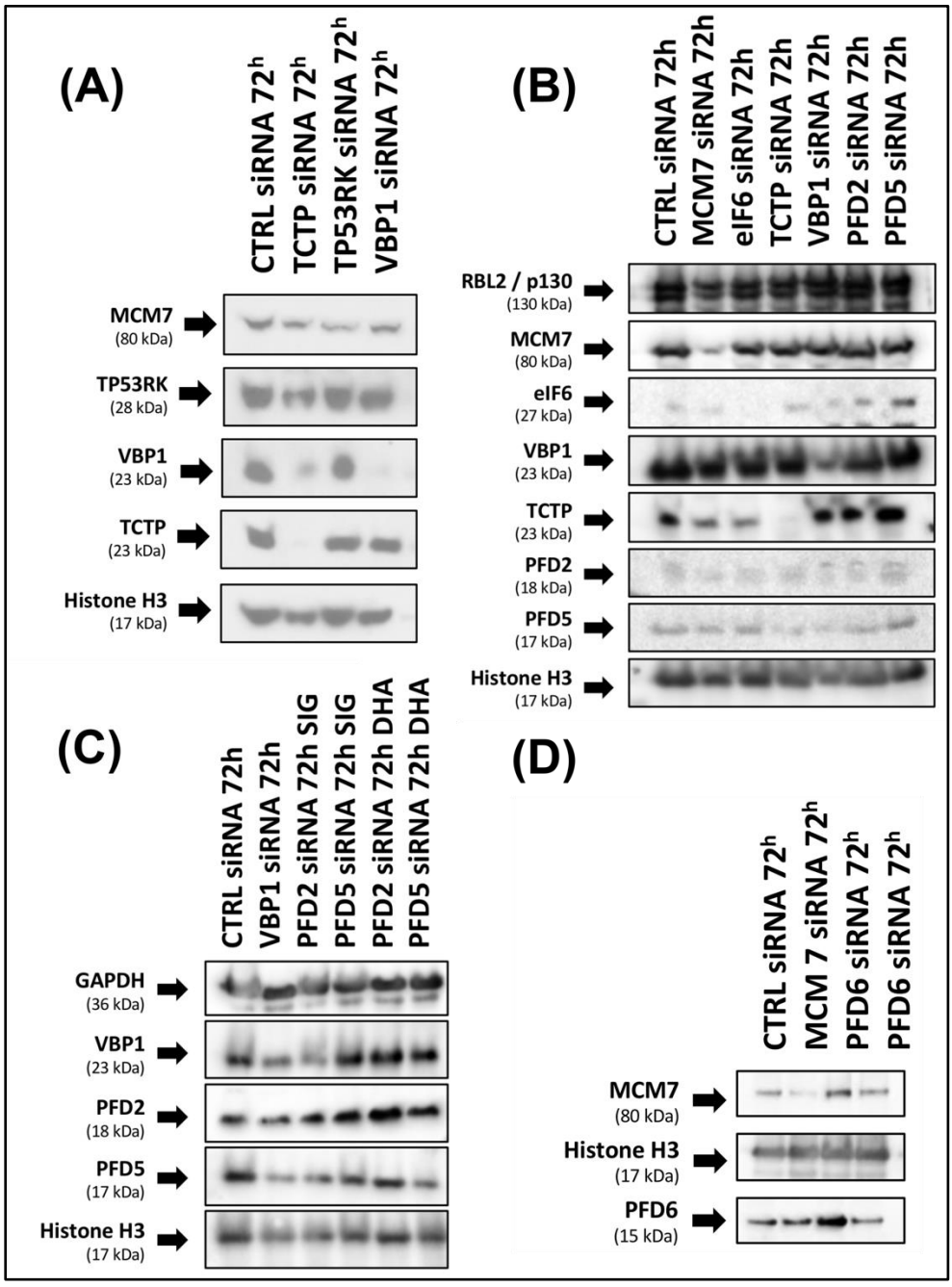


Figure 3.3-9: T-Cell transfections of Prefoldins, VBP1 and TP53RK siRNA.

The siRNA to the targets shown in the Figure were transfected into T-Cells as for Figure 1.3-7. Western blots were carried out for: **(A)** TP53RK, TCTP, and VBP1; **(B)** TCTP, VBP1, PFD2, PFD3, eIF6 and MCM7; **(C)** VBP1, PFD2 and PFD5 and **(D)** PFD6. SMARTpool siRNA were used in each case, except experiments shown in Panel C where "SIG" indicates that custom siRNA was used. DHA: SMARTpool. Blots were also probed for Histone H3, which was used as a loading control.

Based on *Arabidopsis* experiment carried out in parallel (see Chapter 4, Section 4.3.2.6) in which the *PFD6* gene knockout resulted in smaller plants, slower speed of growth and shorter roots, I tested whether PFD6 siRNA could be used to reduce the Pfd6 protein in human T-Cells. The Pfd6 protein is induced by stimulating quiescent T cells with CD3/CD28 (Figure 3.3-9 panel D), however transfecting PFD6 siRNA into quiescent T cells did not reduce the Induction of the protein caused by CD3/CD28 stimulation, either using the standard amount of siRNA used in most experiments in the laboratory (Figure 3.3-9 panel D) or using twice this amount (data not shown). Most siRNA pools used to prevent the induction of many different targets in our laboratory have been effective, but one siRNA pool used to reduce the induction of Cdc6 in another project in the laboratory did not work (Dr Orr, pers. comm.). The possible reason behind the failure to reduce the levels of induction of TP53RK, PFD2, PFD5 or PFD6 could be that the siRNA sequences in the SMARTpool used in these experiments were not effective or that they are degraded too quickly to be effective. This might be remedied by re-transfection with more siRNA after a set time in culture, but this was not tested.

3.3.4 Effects of reduction of individual proteins in siRNA experiments

3.3.4.1 Cell cycle analysis by flow cytometry

In the previous section, I showed experiments in which siRNA could be used to reduce the induction of TCTP, VBP1 and eIF6 in human T-Cells. I determined whether reducing the expression of these proteins affected entry into the cell cycle by flow cytometry. Reducing TCTP, VBP1 or eIF6 did not cause a consistent, reproducible change in the percentages of cells in individual cell cycle phases as compared with T-Cells transfected with control siRNA. The difference in the % of cells in S-phase in (Figure 3.3-10) was not observed in all experiments. In contrast, Mcm7 siRNA reduced the percentage of cells in S-phase, as shown above (see Table 3.3-4).

Table 3.3-4: Distribution of cells in cell cycle phases after transfection with siRNA: TPT1, VBP1, eIF6 and MCM7

Cell Cycle phase	*CTRL*	*TPT1*	*VBP1*	*eIF6*	*MCM7*
Sub G_0	0.25%	0.46%	0.37%	0.36%	0.38%
G_0	36.10%	32.60%	32.80%	37.40%	55.40%
G_1	19.60%	27.40%	28.20%	24.10%	13.00%
S	34.30%	31.80%	30.50%	30.30%	21.60%
G_2-M	9.01%	6.81%	7.05%	6.68%	7.72%

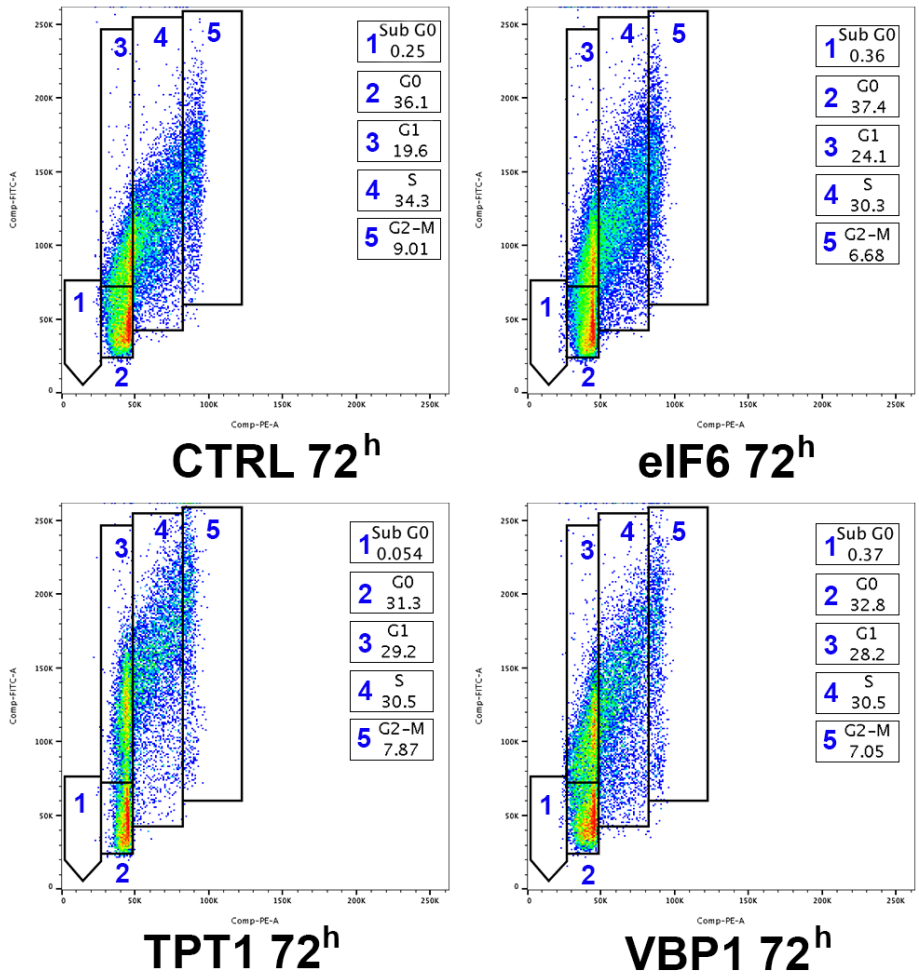


Figure 3.3-10: Cell cycle profile of T-Cells transfected with TPT1, VBP1, and eIF6 siRNA

T-Cell transfections with TPT1, VBP1 and eIF6 siRNA as well as non-targeting control siRNA were carried out and the percentage of cells in each cell cycle phase was determined by flow cytometry of PI and FITC stained cells.

3.3.4.2 Analysing the cell size

3.3.4.2.1 Identifying which siRNA reduces TCTP

In previous work by the laboratory to reduce the levels of Mcm7 and eIF6 using siRNA, each of the four siRNA in the SMARTpool was tested to determine whether more than one was responsible for causing the reduction. In my work, I carried out experiments to test this for the TCTP SMARTpool, which I used in the experiments described above. I determined whether more than one siRNA in the pool caused a reduction of the Tctp protein. Each of the four siRNA was tested by transfection T-Cells, using the experimental protocol used for Figure 3.3-8The levels of the Tctp protein were determined by western blotting and the data are shown in Figure 3.3-11. The same dosage of siRNA was used in all the experiments. The individual TCTP siRNA 05 and 16 both caused a reduction of Tctp protein, whereas siRNA 01 and 06 had less of an effect. Therefore, since the effects of the pool are caused by more than one siRNA, the SMARTpool was used in further experiments.

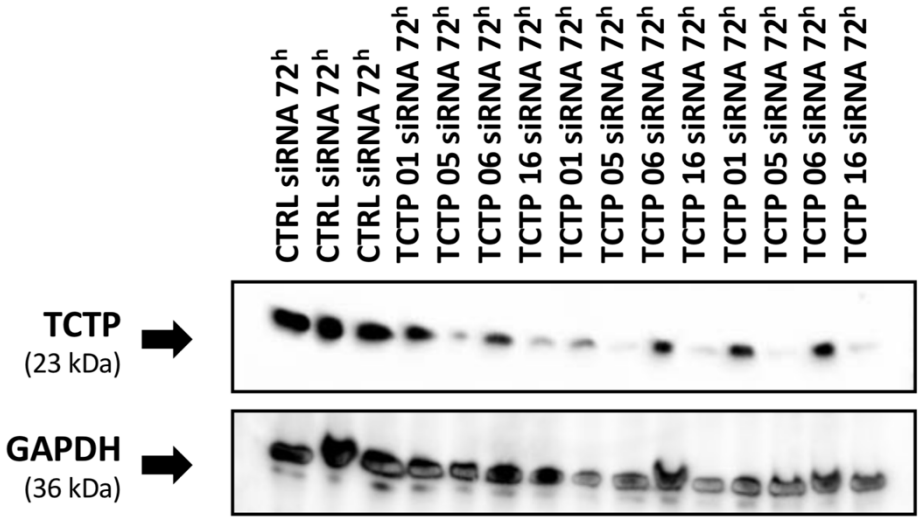


Figure 3.3-11: Efficiency of individual siRNA strands.

T-Cells were transfected with single TCTP siRNAs corresponding to each of the four siRNAs in the SMARTpool used in Figure 3.3-10. These siRNAs are 01, 05, 06 and 16. The experiment was carried out according to the method used for Figure 1.3-13, and the levels of Tctp protein were determined by western blotting. GAPDH was also detected on the blot and is used as the loading control.

3.3.4.2.2 TCTP protein and cell size

I investigated whether reducing the expression of the TCTP protein affects the size of human T-Cells, based on forward scatter data exported from flow cytometry analyses (Figure 3.3-12. In each experiment, the TCTP siRNA caused a reduction of Tctp protein, as determined by western blotting (data not shown). I extracted the mean FSC-A data for T-Cells transfected with control and TCTP siRNA (see Figure 3.3-13). The distribution of cell size is wider in cells transfected with TCTP siRNA as compared with the control. In four of the six independent experiments, cells had higher FSC-A mean values than the control, while in two cases the mean values were smaller (Figure 3.3-14). It is possible that these differences are due to differences in the percentages of T-Cells in cell cycle phases in each experiment. To determine whether this is the case I extracted FSC-A data of individual cell cycle phases (Figure 3.3-15). Furthermore, I added histogram charts for the relative protein content (FITC staining), as compared with the control population of cells (Figure 3.3-16). The data show that there are significant differences (< 0.05) in FSC-A values between control and TCTP knock-down population based on Student's *t*-Test and the cells in G_0/G_1 are significantly smaller. I used the Student's *t*-test to compare the distribution of different populations. Although Kolmogorov-Smirnov (K-S test) also showed that there are significant differences between the control and TCTP data. As these tests show significance, reducing Tctp protein expression in T-Cells entering the cell cycle from quiescence has a statistically significant effect on cell size. However, differences in cell size shown here are small and more experiments need to be done to determine whether this change in size is time-dependent as previous experiments in the laboratory have shown that cells transfected with eIF6 siRNA revert to a normal size at later time-points and earlier time-points might be informative for TCTP siRNA experiments. Also, further experiments are needed to determine whether such differences are biologically significant by investigating the precise cellular mechanism.

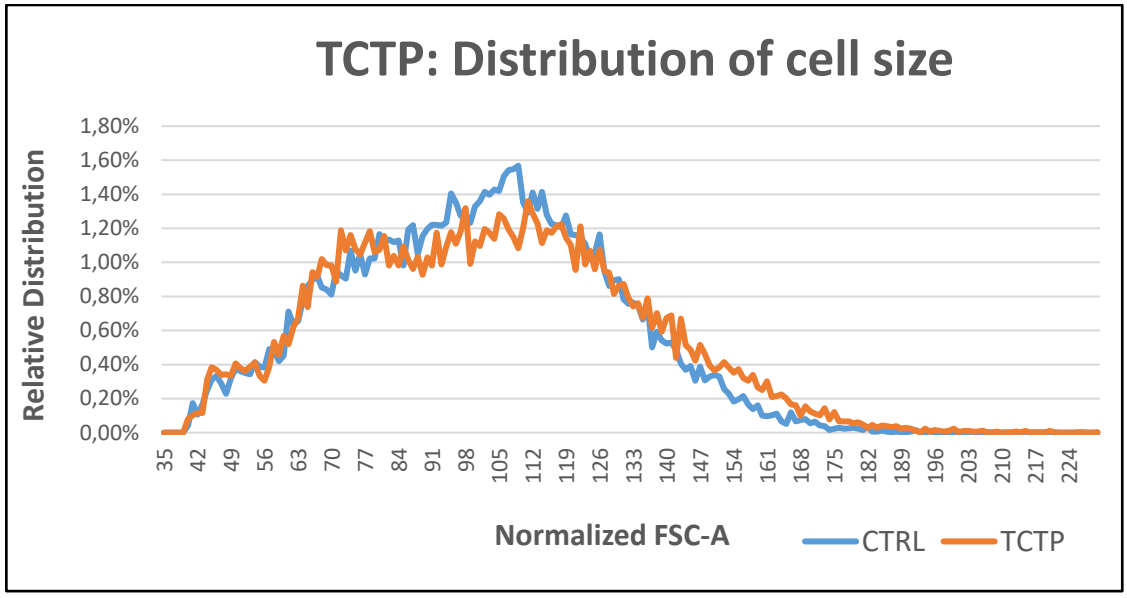


Figure 3.3-12: TCTP: Distribution in the size of single cells.

T-Cells were transfected with control or TCTP siRNA as described for Figure 1.3-13 and the FSC-A values for cells in each population were analysed by flow cytometry. This distribution is a normalization result of the six independent experiments.

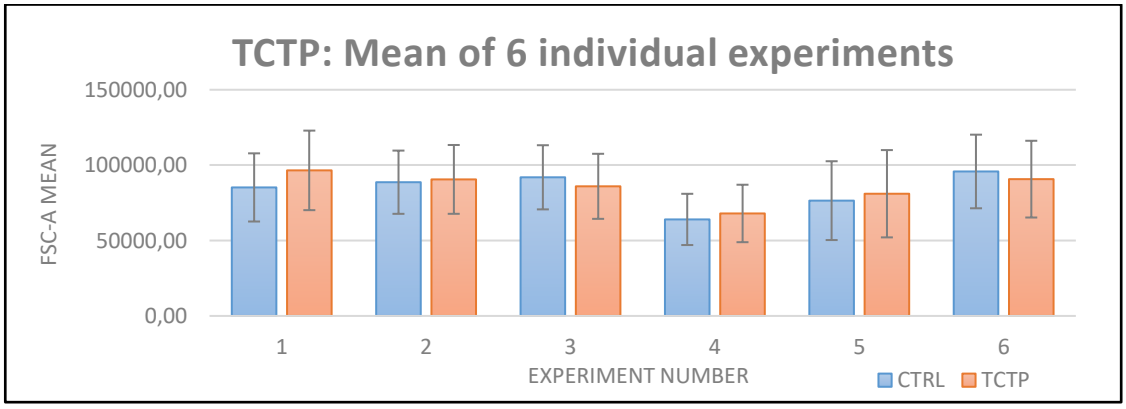
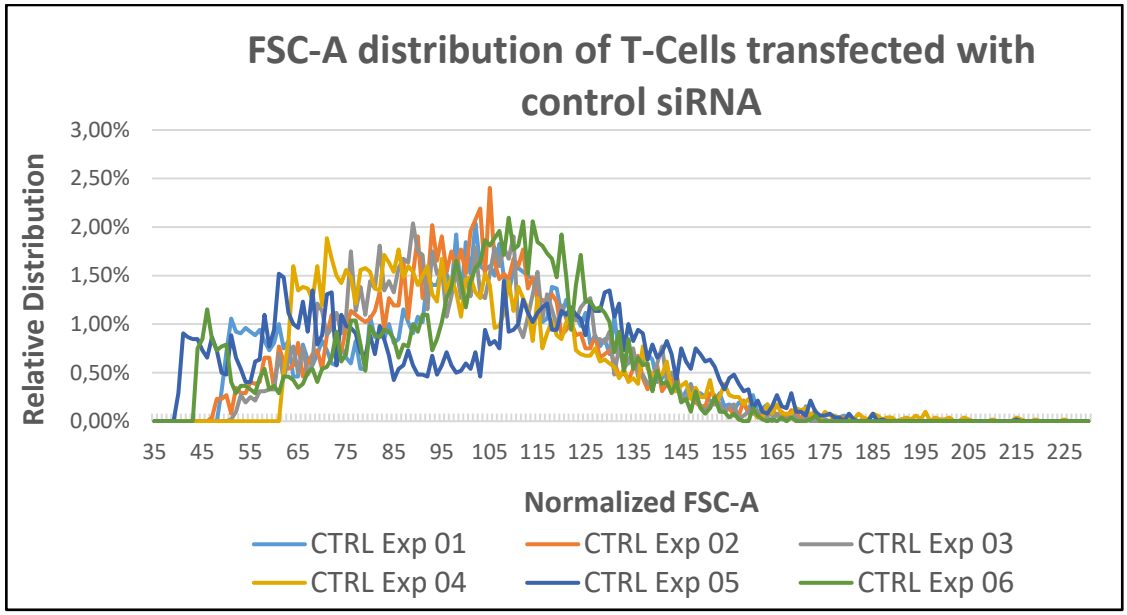


Figure 3.3-13: TCTP: Comparing mean values for six different, individual experiments.

The figure shows the mean of the FSC-A data from six individual siRNA experiments in which T-Cells were transfected with control or TCTP siRNA.



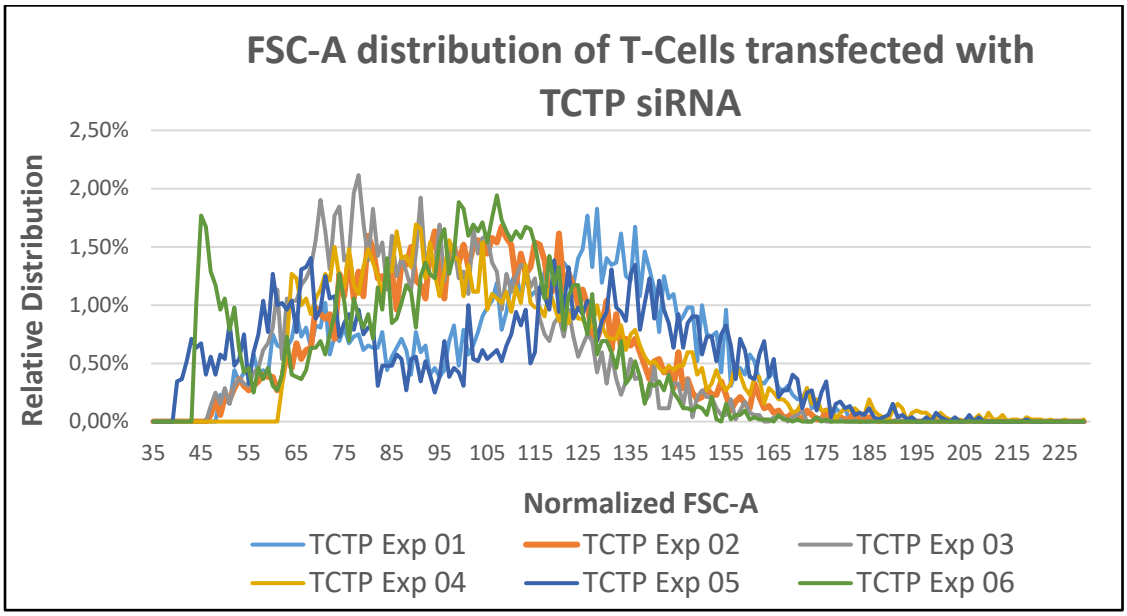


Figure 3.3-14: Comparing distributions of control cells and cells transfected with TCTP siRNA

FSC-A values for single T-Cells from each of the six TCTP and control siRNA experiments described in the text are shown.

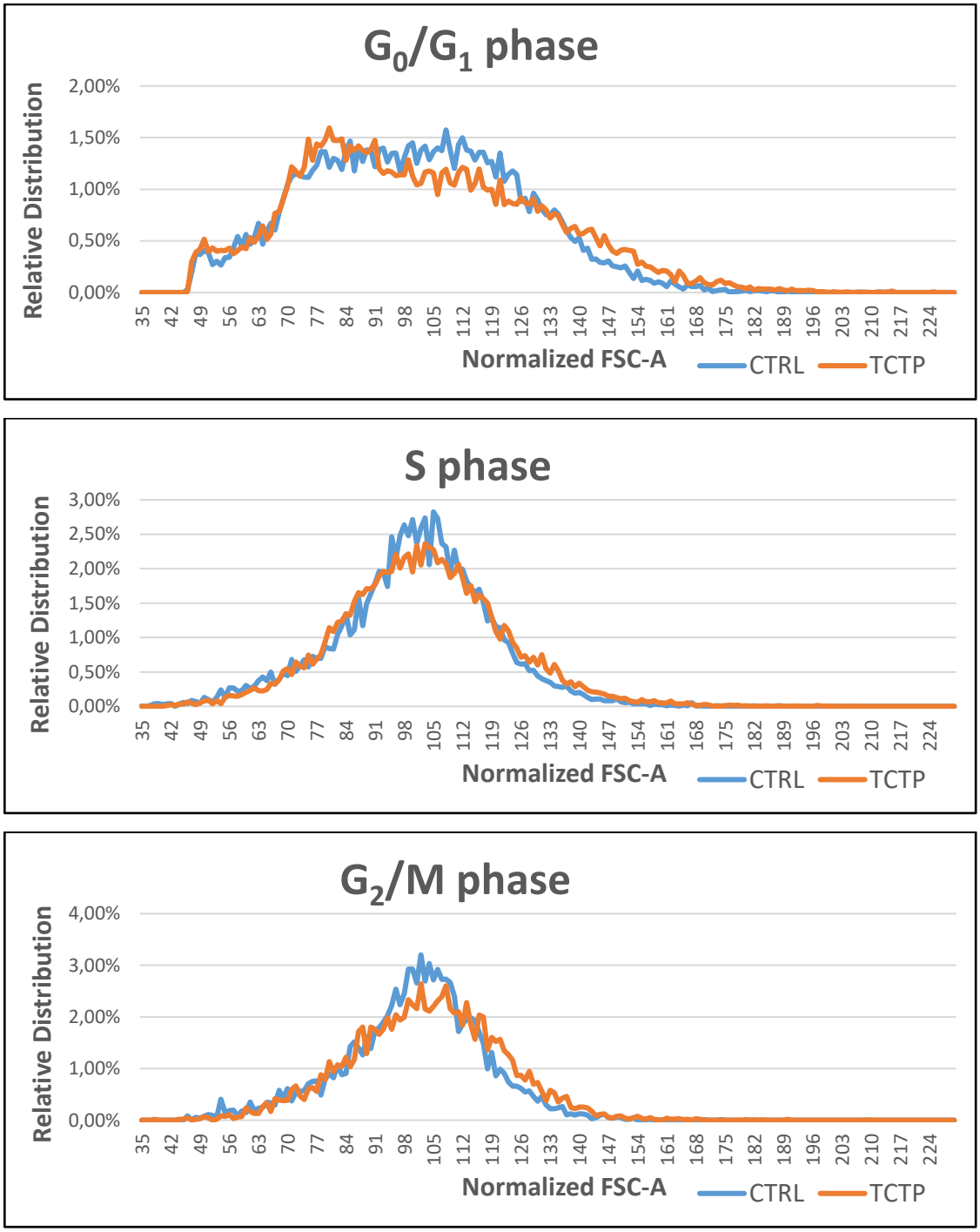


Figure 3.3-15: TCTP and control siRNA: Distribution of cell sizes in individual cell cycle phases.

See text for details. These results are based on the normalization of FSC-A values for six independent siRNA experiments.

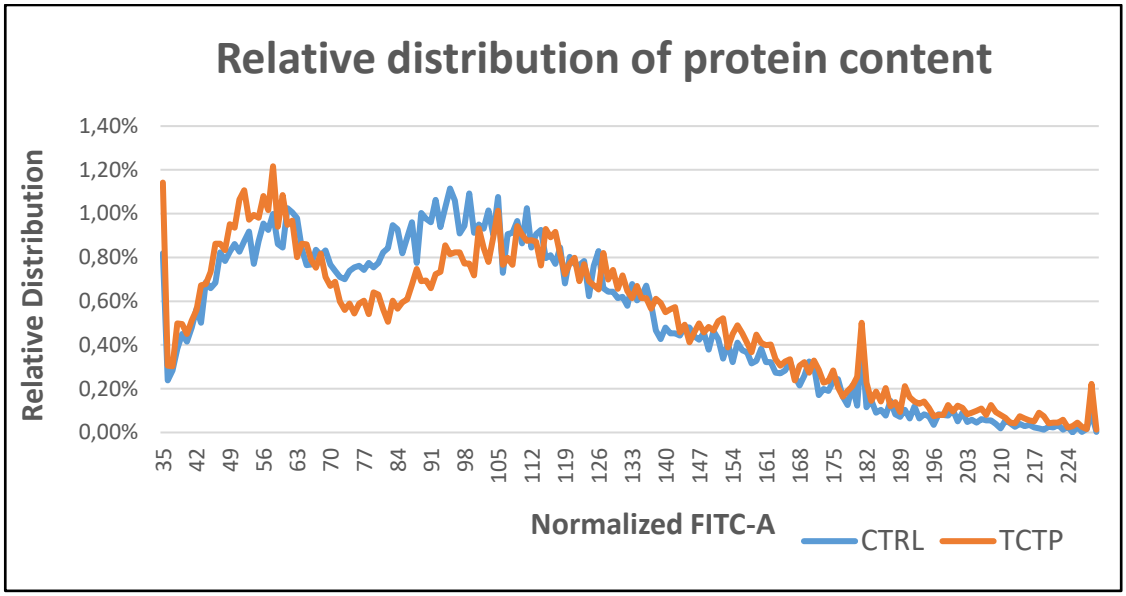


Figure 3.3-16: TCTP and control siRNA: Relative protein content

FITC-A intensity values, which represents the protein content, for single T-Cells from six TCTP and control siRNA experiments.

3.3.4.2.3 VBP1 protein and cell size

VBP1 is a component of the Prefoldin complex and I was able to reduce the expression of Vbp1 by transfecting siRNA into T-Cells (see Figure 3.3-17). First, I analysed whether reducing the levels of Vbp1 affects the FSC-A values for single cells (Figure 3.3-17), or for cells in individual cell cycle phases (Figure 3.3-18). The data show that reducing the expression of Vbp1 does not significantly affect the FSC-A values of T-Cells in any cell cycle phase. Furthermore, I analysed the protein content (FITC-A values) of these experiments and the relative distribution is shown in Figure 3.3-19. The FSC-A data obtained for the Vbp1 knockdown experiments are not statistically different from the control based on the Student's *t*-test, indicating that reducing Vbp1 expression does not affect the size of human T-Cells.

However, there is an apparent difference in the FITC-A (cellular protein content) values for T-Cells transfected with control and VBP1 siRNA (Figure 3.3-19). The reason behind this could be that VBP1 knock-down changes the cellular protein expression levels but cell size Is not significantly affected. Given that FSC-A and FITC-A parameters agree under normal conditions post CD3/CD28 stimulation (Orr et al., 2012) further experiments are now required to determine whether cell size and protein content can be dissociated by reducing Vbp1 expression with siRNA.

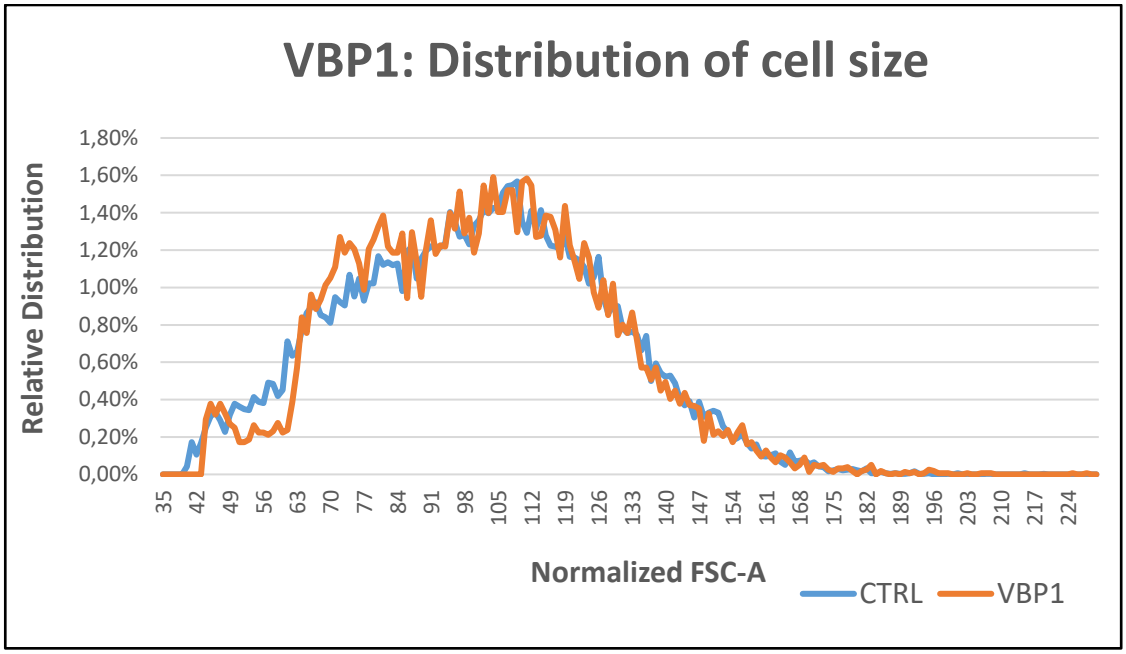


Figure 3.3-17: T-Cells transfected with VBP1 or control siRNA: Distribution of cell size

See text for the details. These results are based on the normalization of FSC-A values of six independent experiment. The normalized distribution of cell size showed no difference between the control and VBP1 knock-down experiments.

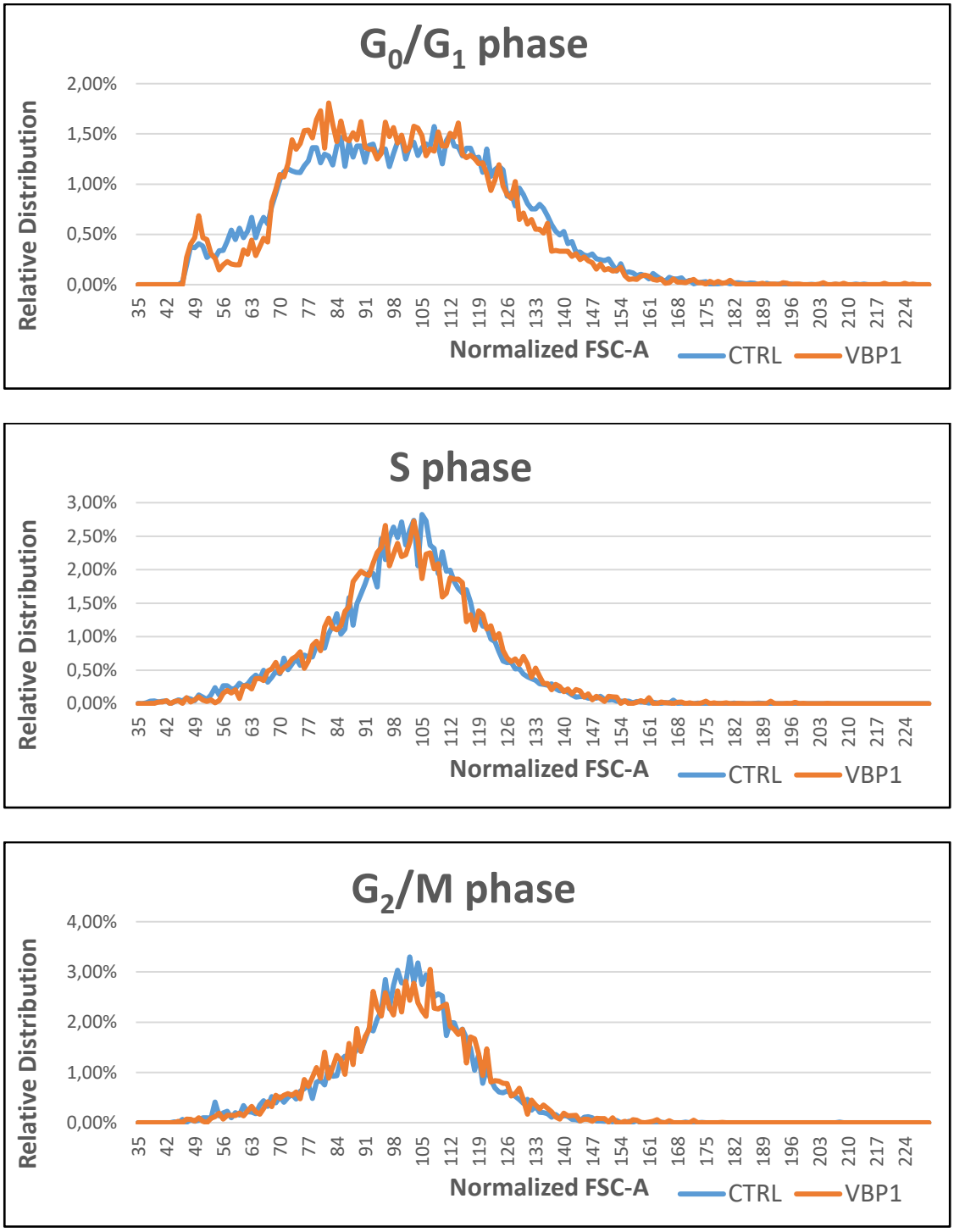


Figure 3.3-18: T-Cells transfected with VBP1 or control siRNA: Distribution of cell size in individual cell cycle phases

See text for the details. These results are based on the normalization of six independent experiment

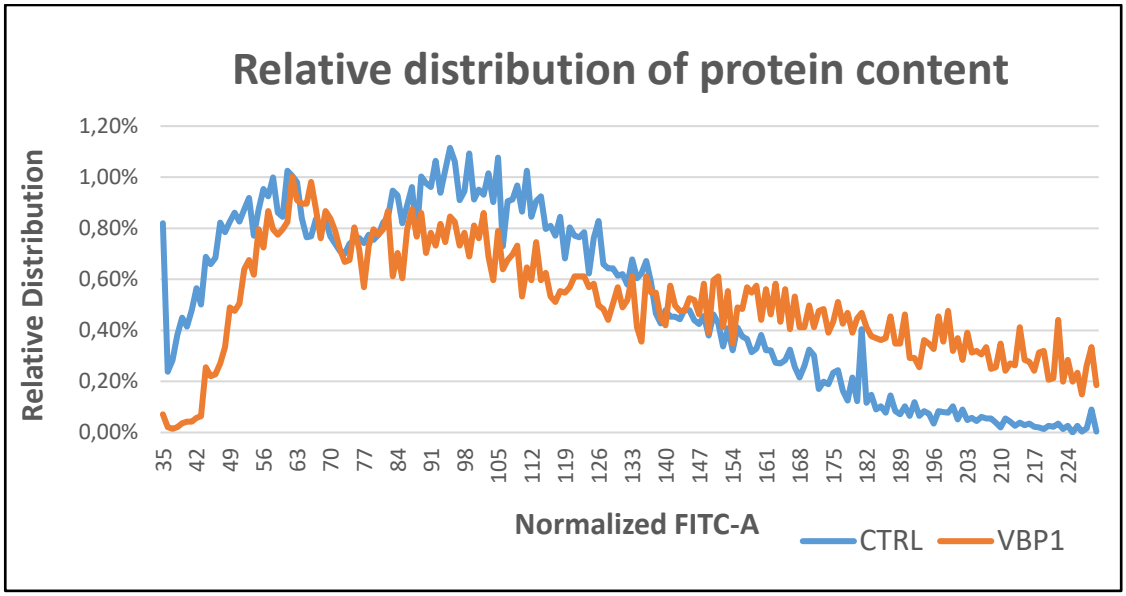


Figure 3.3-19: VBP1 and control siRNA: Relative protein content

FITC-A intensity values, which represents the protein content, for single T-Cells from the four VBP1 and control siRNA experiments.

3.3.4.2.4 MCM7 protein and cell size

I always used MCM7 siRNA as a technical control for each transfection experiment, and I also determined the FSC-A profiles recorded by flow cytometry. FSC-A or other measures of cell size had not been analysed in previous work published by my laboratory (Orr et al., 2010). Figure 3.3-20 shows normalised data of 15 individual, independent experiments using T-Cells Isolated from different donors. 9 of these experiments were carried out previously by Dr Stephen Orr for work published in (Orr et al., 2010). The data shows that the FSC-A size distribution of cells transfected with MCM7 siRNA, in which Mcm7 protein is reduced is smaller. The data are quantified in Figure 3.3-21. I found that FSC-A mean values of individual cells are statistically lower in T-Cells transfected with MCM7 siRNA as compared with non-targeting control siRNA, except for two experiments (#12 and #13). A graph of the relative protein content (FITC-A staining) of cells transfected with Mcm7 siRNA was also statistically different from the control (Figure 3.3-23.).

Because of the difference observed in FSC-A values, I also investigated the FSC-A of cells in individual cell cycle phases based on the gating I used in flow cytometry analyses (Figure 3.3-22). I found that in G_0/G_1 phases, cells transfected with MCM7 siRNA have smaller FSC-A values than the control and the distribution of FSC-A values is sharper. The distribution range in S phase is much wider and the mean FSC-A is higher when Mcm7 is reduced than in control-siRNA transfected cells. Cells in G_2/M have FSC-A values that are similar to the control. These data show that reducing Mcm7 causes a reduction in the FSC-A of cells in G_0/G_1 and an increase in S-phase. Alterations in FSC-A distributions are consistent with the expression of Mcm7 affecting cell size of human T-Cells. The biological relevance of this results is while the MCM complex is responsible for initiating DNA replication in late G_1 phase, "licensing" cells to enter S-phase (Blow & Hodgson, 2002; Oehlmann et al., 2004) it means that inhibiting one of the members can also significantly reduce cell size. However, MCM proteins also form complexes with RNA-Polymerase II and may have a role in transcription (Holland et al., 2002; Yankulov et al., 1999). Reducing Mcm7 may therefore affect the expression of specific genes and affect cell size Indirectly.

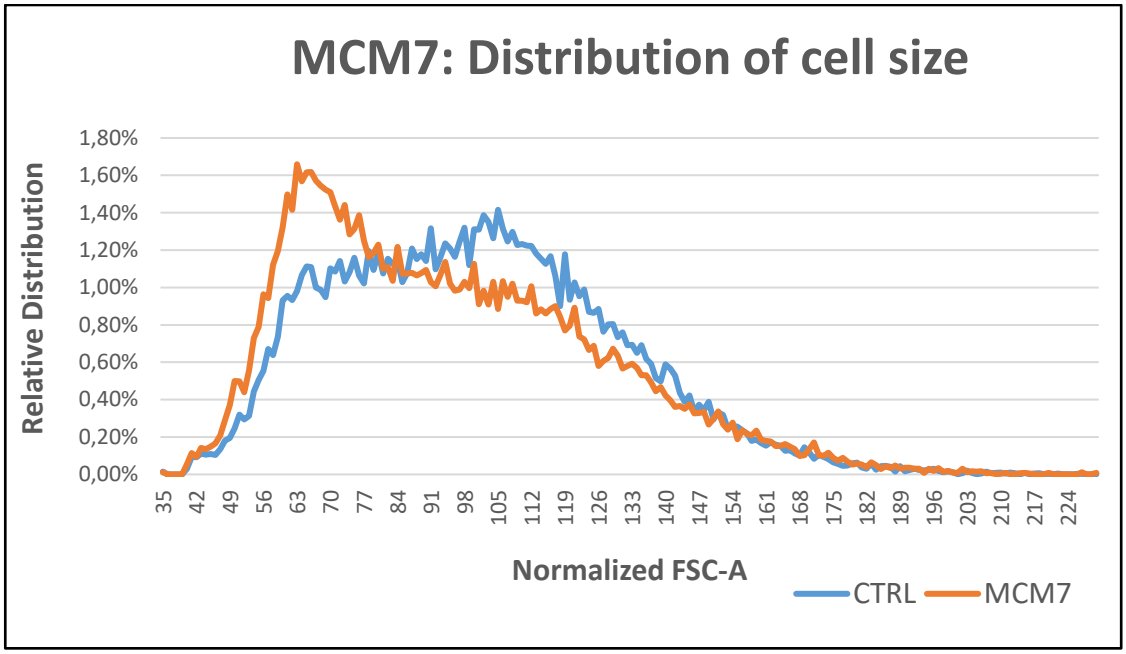


Figure 3.3-20: T-Cells transfected with MCM7 or control siRNA: Distribution of cell size in single T-Cells

See the text for details. This distribution is a normalization of FSC-A values from of 15 independent experiments.

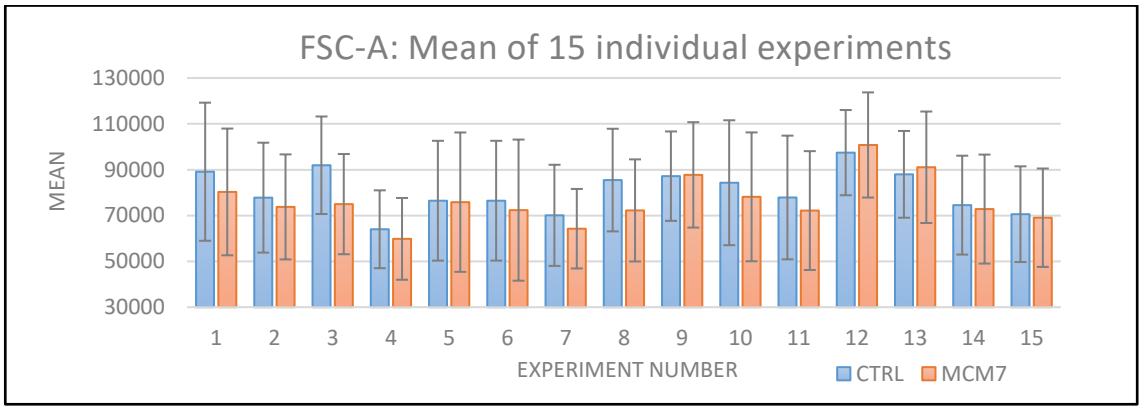


Figure 3.3-21: MCM7 and control siRNA transfected T-Cells: Comparing mean FSC-A values of 15 different T-Cell experiments

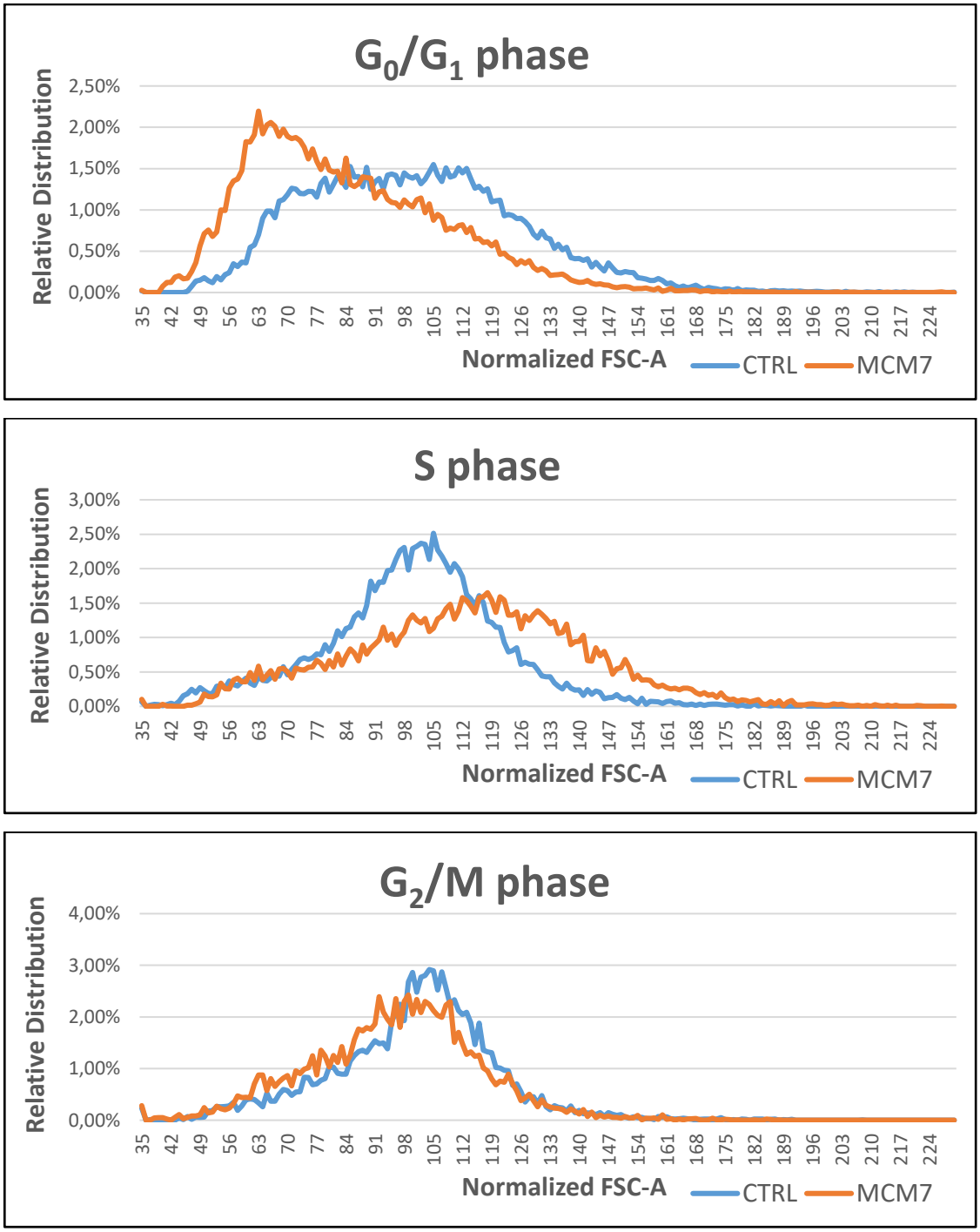


Figure 3.3-22: T-Cells transfected with MCM7 or control siRNA: Distribution of cell size in individual cell cycle phases

See text for the details. These results are based on the normalization of 15 independent experiment

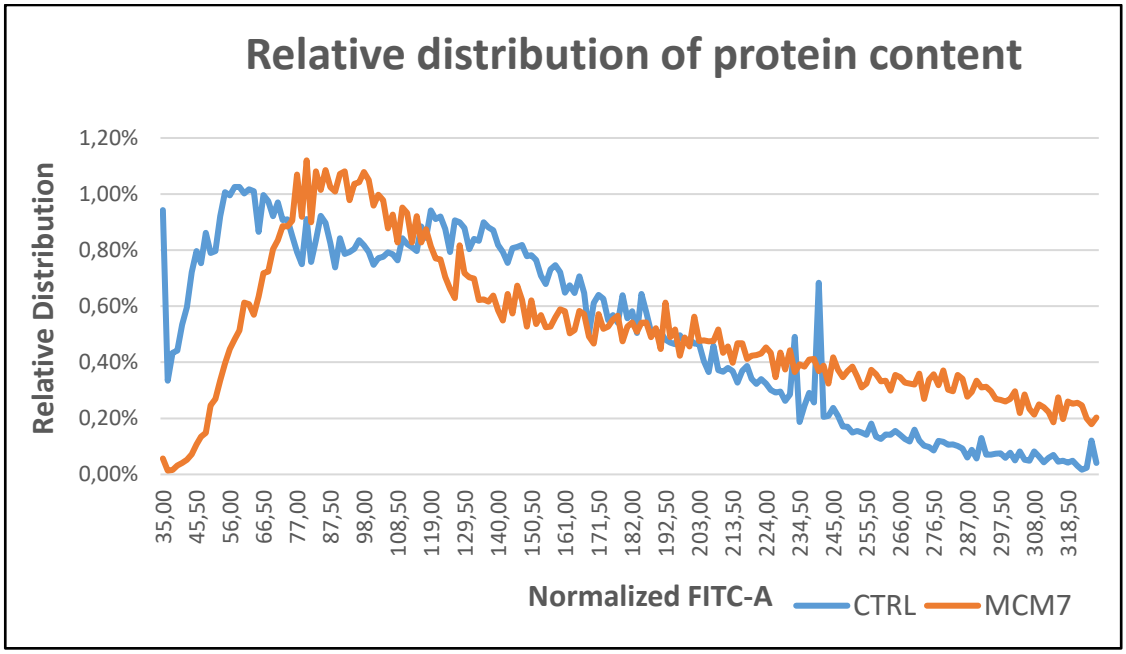


Figure 3.3-23: MCM7 and control siRNA: Relative protein content

FITC-A intensity values, which represents the protein content, for single T-Cells from 15 individual MCM7 and control siRNA experiments.

3.4 Discussion

3.4.1 T-Cell- entry to the cell cycle

In this chapter, I described experiments I carried out using human peripheral blood T-Cells. I showed that non-stimulated, quiescent T-Cells could enter the cell cycle after mitogenic stimulation *via* CD3/CD28 cell surface receptors. My results clearly show that T-Cells enter and progress through the cell cycle with kinetics similar to that published previously by our laboratory (Lea, Orr, et al., 2003), double their DNA content (judged by PI staining) and increase their protein content (judged by FITC staining) as described in (Thomas, 2004). Moreover, I showed that the pRB protein is hypo-phosphorylated in quiescent cells and is phosphorylated on serine 807 and serine 811 in response to CD3/CD28 stimulation, consistent with the cell cycle transition from G_0 to G_1 phase (Lea, Orr, et al., 2003; Ren & Rollins, 2004). The distribution of cells in different cell cycle phases are also in line with the literature as 99% of the T-Cells did not enter S phase until 24-48h post CD3/CD28 stimulation. This agrees with the published literature as CFSE[8] staining experiments carried out by my laboratory showed that by 24h post stimulation 1% of cells divide once, by 48h this increases to 16.4%, and finally 72h post stimulation 36% of cells have divided once and 1% twice (Lea, Orr, et al., 2003).

The cell size of the human T-Cells is conserved between 7 and 10 μm in diameter. Predominantly memory T-Cells have a strict characteristic in size. T-Cells follows a quantititative regime of cell size as they experience CD3/CD28 stimulation: they double they size during the mitosis then halves after the division.

[8] *Carboxyfluorescein succinimidyl ester* is a cell staining dye, that is used to label and track cell proliferation. The CFSE per cell halves as a consequence of each cell division.

3.4.1.1 Using flow cytometry to measure cell size

The method to measure cell properties such as cell size and volume by flow cytometry is commonly used in the literature (Brown & Wittwer, 2000). A parameter known as forward scatter area (FSC-A) is usually used to determine cell size, but a recent study also reported the use of FSC-W and side scatter area (SSC-A) (Tzur et al., 2011). Studies by my laboratory have used FSC-A and protein content (FITC staining) to measure cell size (Orr et al., 2012) and a recent study also confirmed that there is a high correlation between forward scatter area and cell size (Miettinen & Björklund, 2016). I tested whether the FSC-A and FITC-A of T-Cells stimulated with CD3/CD28 changes with time post CD3/CD28 stimulation. My data show that this does occur and is consistent with the increase in cell size known to occur as cells progress through the cell cycle.

Studies by other research groups have also confirmed that using forward scatter data to measure cell size is an appropriate method. Researchers used flow cytometry data to measure cell size of human stem cells (Machado et al., 2013), human epithelial cells (De Paiva et al., 2006) and other organisms such as *E. coli* (López-Amorós et al., 1994). While there are many different flow cytometry machines available in my department and others, others have shown that there could be a significant difference between machines (Standerholen et al., 2014). Therefore, it is important to use the same machine for a study, which is what I was able to do for the work presented in this Thesis. The way in which electronic "gates" are set and how experiments are carried out and analysed can affect the data obtained. To reduce experimental variability, all of the human T-Cell experiments presented in this thesis were carried-out by me using the same FACS machine and I used the same techniques and the protocols that my laboratory published previously (Lea, Orr, et al., 2003). I ensured technical reproducibility by using positive controls in each experiment and electronic compensation and setting electronic gates were according to standard methods used by the laboratory.

3.4.1.2 Using a Cellometer to measure cell size

It has been reported that a Cellometer is a good way to measure cell size, which has been used in assays of adipocytes (Lee et al., 2012), stem cells (Lo Surdo & Bauer, 2012) and cells producing recombinant adeno-associated virus (Cecchini et al., 2011). I measured the size of T-Cells using Nexcelom slides and the automated Cellometer. The T-Cells were in a quiescent state before I added CD3/CD28 beads to stimulate them to enter and progress through the cell cycle. However, cell size measurement data of T-Cells generated by the Cellometer were inconsistent and unreliable and did not increase with time after CD3/CD28 stimulation. The expression of cellular adhesion molecules on the surface of T-Cells increases in response to CD3/CD28 stimulation, which results in the formation of large clumps of cells that are difficult to disaggregate completely. The inconsistency in the values determined by the Cellometer may be caused by it not being able to distinguish individual T-Cells in clumps. Because of the inconsistency and the fact that the Cellometer may exclude the CD3/CD28-activated, larger cells, which are the cells responding to mitogenic stimulation, I did not use the Cellometer to analyse cell size in my siRNA transfection experiments.

3.4.2 siRNA experiments- TP53RK and Prefoldins

As I showed in Results section 3.3.3.2 , I was not able to reduce the expression of TP53RK or the Prefoldin proteins PFD2, PFD5 or PFD6 using SmartPools of four siRNA. I ordered custom single siRNA sequences for PFD2 and PFD5 based on (Abe et al., 2013) and I tested different antibodies recognising epitopes in different domains of the proteins. None of the siRNAs tested caused a reduction in the expression of the protein encoded by the target mRNA, so I excluded TP53RK, PFD2 and PFD5 from further analysis. A PFD6 siRNA SMARTpool was tested late during my PhD and therefore I was not able to test custom siRNA or other antibodies. It is not clear why the siRNA to each of these targets did not cause a reduction in protein expression since MCM7 siRNA, which was used as a technical control for the Nucleofection method in most experiments, caused a reduction of Mcm7 protein. Only a few siRNA experiments in the laboratory have been unsuccessful, notably CDC6 siRNA (unpublished). Further

work into TP53RK and Prefoldin mRNA expression and turnover in T-Cells would need to be done before further siRNAs to these targets are tried. Experiments in the laboratory have also shown that certain siRNAs are degraded more quickly than others. For example, *ΔTP73* siRNA was degraded in T-Cells from 48h post CD3/CD28 stimulation and the Δp73 protein increased to control levels after that (Chronis *et al.* - in preparation). Therefore, TP53RK and Prefoldin siRNA turnover in T-Cells may also be a factor, and different siRNA sequences and other methods (see Chapter 5) would have to be investigated.

3.4.2.1.1 VBP1 and cell size

VBP1 is one of the two alpha-subunits of the prefoldin complex (Martín-Benito et al., 2007). The data showed that unlike in other organisms reducing the expression of VBP1 has no effect on the size or cell cycle entry of the T-Cells analysed in my experiments. I repeated experiments more than six times, and there were no significant differences. Many of the other prefoldin complex members have been reported to regulate cell size in other organisms (Gu et al., 2008; Jorgensen et al., 2002; Rodríguez-Milla & Salinas, 2009). In human T-Cells, the reduction of VBP1 may not affect cell size because of functional compensation. There are many examples of functional compensation in human cells by members of a protein family. Examples of this for proteins involved in the cell cycle are the CDKs, apart from Cdc2/CDK1 (Malumbres & Barbacid, 2009). The *RBL2* (encodes p130) knockout is compensated by other members of the pRb in murine T-Cells (Mulligan et al., 1998). Compensation has also been shown to occur in cancer cells in which somatic mutations inactivate specific genes with no functional effect (Cereda et al., 2016). It is possible that the reduction of Vbp1 in human T-Cells is compensated by other members of the family. In another project in the laboratory, the Prefoldin protein complex has been isolated by biochemical fractionation and in-line mass spectrometry analyses of the proteins present (Orr *et al.* - in preparation). It would be interesting to determine the effect of reducing Vbp1 on components of this complex and protein folding (Sahlan et al., 2018).

3.4.2.2 TCTP

I showed that reducing TCTP in human T-Cells with siRNA has an effect on cell size. However more experiments need to investigate the underlying mechanisms and the biological significance and whether the mechanism directly or indirectly has an effect on cell size. TCTP is a conserved protein in many species (Hinojosa-Moya et al., 2008) and it has been associated with many different functions in human cells, including cell stress responses, anti-apoptotic activity, promoting cell growth and division moreover, it is also released from immune cells and has a role in cytokine release (Bommer, 2017). TCTP regulates cell size in *Drosophila melanogaster* (Hsu et al., 2007) and it is important for regulating the development of organ size and growth (Hong & Choi, 2013). Moreover, the gene knock-out has been reported to reduce cell size in *Arabidopsis* (Berkowitz et al., 2008). These effects could be because the TCTP protein has a microtubule-stabilising activity in a cell cycle-dependent manner (Gachet et al., 1999). Finally, as the name suggests, the protein plays a role in the development of carcinomas (Gross et al., 1989). The functions of TCTP in apoptosis and cell proliferation are important in this context, but it may also be involved in invasion and metastasis during the later stages of cancer progression (Bommer, 2017).

3.4.3 MCM7 results

Mcm7 is a member of the Minichromosome Maintenance Complex (MCM), which forms the DNA helicase is essential for replication during S-phase (Zhai & Tye, 2017; Zhai et al., 2017). It was shown previously by my laboratory that reducing Mcm7 (or Mcm4) levels in T-Cells during cell cycle entry from quiescence caused genomic Instability, DNA damage and affected cell cycle progression (Orr et al., 2010). Because of this previous work, I used siRNA to reduce Mcm7 expression in T-Cells as a technical control for my studies of other proteins. However, the effect of reducing Mcm7 on cell size was not investigated in work by Orr *et al.* and so I analysed the effect of reducing Mcm7 alongside my other experimental samples. Analyses of T-Cell size in different cell cycle phases in my experiments showed that the sizes of cells in G_0/G1 is reduced and increases in S-phase when the Mcm7 expression is reduced with siRNA. To

determine whether my results agree with those produced earlier by Dr Orr, I obtained flow cytometry data from him for nine experiments use in (Orr et al., 2010). Analysis of his data confirmed my results. The effect of reducing Mcm7 on cell size could either be a specific effect of Mcm7 or as a component of the MCM complex. Therefore, I also carried out a limited number of siRNA knock-down experiments for Mcm4. Based on two individual experiments, reducing Mcm4 also affected cell size in a similar manner, which suggests a role for Mcm7, Mcm4 and possibly the whole MCM complex in the regulation of cell size. Further replicate experiments would need to be done to investigate this, including reducing each of the MCM proteins individually using siRNA used previously in the laboratory (Orr et al., 2010). The increase of cell size in S-phase cause by reducing Mcm7 could be due to the effects on DNA replication, including fork collapse and subsequent DNA repair increasing the time cells take to complete S-phase (*ibid*). However, this does not account for the reduction of cell size in G_1. The MCM proteins also bind to RNA Polymerase II and have roles in other functions, including transcription (Forsburg, 2004). It is possible that reducing Mcm7 (or Mcm4) affects the transcription of genes that are required for cell growth. To this end, I isolated RNA from T-Cells transfected with siRNA before they entered S-phase. These RNAs will be sequenced to determine whether reducing the induction of Mcm7 during G_1 affects the transcription of specific genes in T-Cells and if these are likely to affect processes involved in growth, such as ribosome biogenesis.

3.4.4 Conclusions

The results that I obtained during my PhD suggest that the expression level of TCTP in T-Cells could be a possible regulator of cell size. The Tctp protein is a multifunctional protein that has several distinct functions including involvement in apoptosis and as a chaperonin protein and an effect on cell size could be indirectly due to one or more of these mechanisms.

I have also shown that reducing the expression of Mcm7, a technical control in my experiments, affects cell size, which could be due to a role in DNA replication and/or transcription.

In summary, I have identified two proteins that are possibly important for regulating cell size of human T-Cells in specific cell cycle phases. More experiments need to be done to truly confirm these proteins as cell size regulators and the precise biological mechanisms involved.

Chapter 4

Arabidopsis thaliana

4.1 Introduction

4.1.1 General Background of *Arabidopsis thaliana*

Arabidopsis thaliana is a flowering plant small in size that is a widely used as a model organism in plant biology. Arabidopsis is a multicellular eukaryote plant with a relatively small genome of about 135 megabase pairs (Mbp) with 5 chromosomes that were fully sequenced in 2000 (The Arabidopsis Genome Initiative, 2000).

4.1.1.1 Biology of Arabidopsis

Arabidopsis thaliana plant is a member of the mustard family (Brassicaceae or Cruciferae) and can be found in landsides in the Northern hemisphere including most of Europe, Asia and North America. There are several ecotypes growing in nature, but only two wild-type accessions are used in scientific research: Columbia (or *Col-0*) and Landsberg (Ler) ecotypes. The Figure 4.1-1 shows for a *Col-0* plant, marked with the basic features at 14 days post germination (dpg).

Arabidopsis plant is a relatively short-lived plant with a life cycle of six weeks from seed to seed. It starts with seed germination, then the formation of rosette leaves, followed by the bolting of the main stem, then it starts to flower, and the seeds mature inside the seed pod called silique. The average size of a plant is between 10 to 30 centimetres, while the diameter of rosette leaves is between 2 to 10 cm, both depending on growth conditions. Flowers are usually around 2 cm long, and the sexual reproduction occurs by self-pollination. Leaves are covered with small hair structure called trichomes. Leaves come in a specific order following a growth in spiral that can be counted by selecting the oldest leaf (leaf #1), then counting clockwise towards the right direction. The plants are starting to bolt about 3 weeks after germination then they follow linear growth diagram. A mature plant produces around 5000 seeds in total. Roots have a simple structure with a main root growing following gravity and several lateral branches that emerge laterally (reviewed in Meinke, 1998).

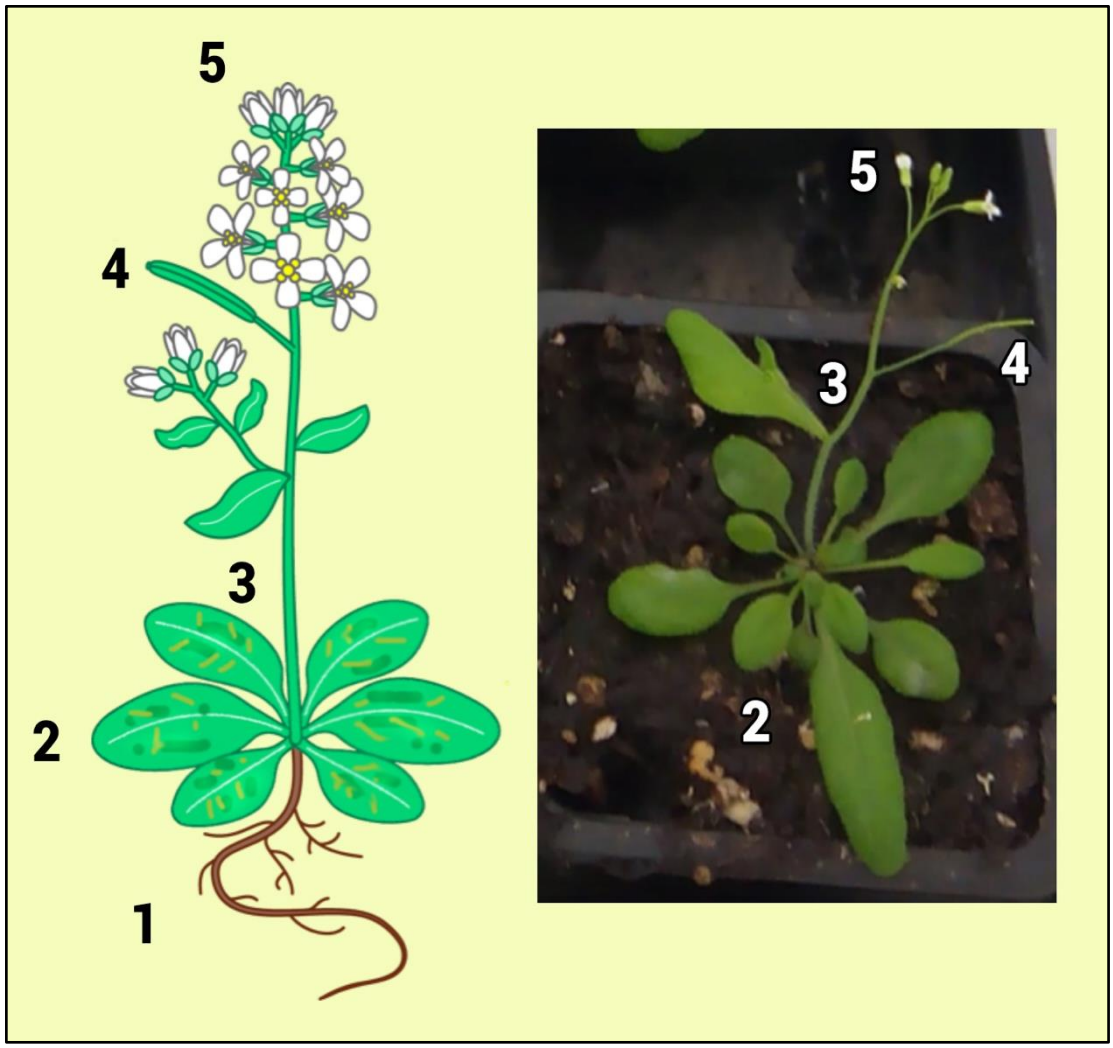


Figure 4.1-1: Schematic picture of Arabidopsis plant

On the left of the figure, a schematic drawing of an Arabidopsis plant can be seen in line with a picture taken 14 days post germination (dpg). On this figure I name four major part of the Arabidopsis plant, that can be experimentally measured:

(1) roots
(2) rosette leaves
(3) flower ramp
(4) siliques and
(5) flowers

4.1.1.2 Use as a model organism

Arabidopsis plant is a model plant of choice allowing the researchers to examine all aspects of plant development and physiology while enabling the use of sophisticated genomic approaches thanks to the small genome. Furthermore, the research using Arabidopsis leads to breakthrough discoveries that are applied to important crops (Meinke, 1998; Provart et al., 2016). *Arabidopsis thaliana* research is important to understand the role of genes using forward or reverse genetics by the relatively easy generation of knock-out or knock-down plant mutants. As of 2014, only 12% of the genes have an experimentally assigned function in Arabidopsis. The vast majority of assigned functions to other gene are potential and based only on sequence similarity assignments (Rhee & Mutwil, 2014).

4.1.1.2.1 T-DNA mutants

A method emerged in the late 1990s using genome-wide random T-DNA insertions to disrupt gene function (O'Malley et al., 2015). This was achieved using a soil bacteria called *Agrobacterium tumefaciens* to integrate a transfer plasmid called T-DNA in the coding regions, promoters or UTRs of genes in the *Arabidopsis thaliana* genome (O'Malley & Ecker, 2010; Tzfira et al., 2004). The most used protocol to generate T-DNA mutants is using the floral dip protocol that consists in simply dipping the flowers into an Agrobacterium solution (Clough & Bent, 1998). In this way, mutant seed banks were created and open to Arabidopsis researchers. There are two large repositories: The European Arabidopsis Stock Centre (NASC: `http://arabidopsis.info/` and the Arabidopsis Biological Resource Center (ABRC: `https://abrc.osu.edu/`). There are also bioinformatic tools (`http://signal.salk.edu/cgi-bin/tdnaexpress`) that help the researchers to visualise the sequence spanning the T-DNA insertions and to further design the PCR primers to check for homozygous plants (O'Malley et al., 2015).

4.1.2 Size Control in Arabidopsis

Over millions of years even closely related species have adopted specific, hugely different organ sizes (Mizukami, 2001) that suggest there are specific size control mechanisms taking place in plants (Czesnick & Lenhard, 2015). The development of plant starts with the growing embryo where basic structures of the plant are established. Later the shoot apical meristem (SAM) is formed, this makes the basis of further development of several structures like leaves and flowers. The SAM cells are very similar and functionally equivalent to stem cells in animals, while it consists of pluripotent cells (Aichinger et al., 2012). These cells can grow and specifies in different structures they form primordiums for organs such as leaf primordium for organ development (Tsukaya, 2013). After forming primordium cells, expansion begins, it is primarily driven by cell-wall modifications accompanied by increasing turgor pressure because of water uptake into the vacuoles (Schopfer, 2006). Several ways affecting and controlling the development of individual leaves, but the final size of a leaf is critically influenced by the duration of cell proliferation (Czesnick & Lenhard, 2015). For the first part, the primordium cell number increases through division, then they start to arrest at the distal tip (Donnelly et al., 1999). This arrest continues and are controlled by several regulator molecules as it can be seen on Figure 4.1-2. These factors are either control the proliferation or the duration of cell expansion to reach the tightly regulated final size of a leaf.

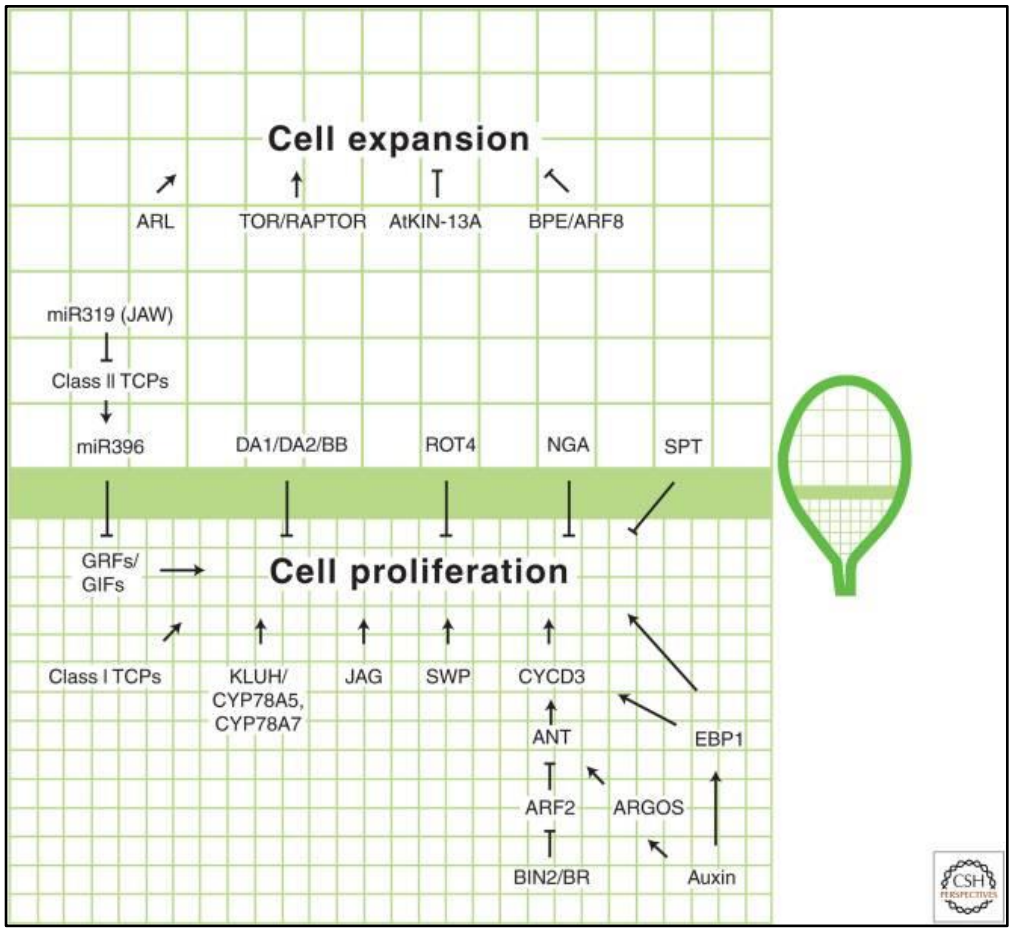


Figure 4.1-2: Main genetic factors that are controlling the growth of plants

The duration of cell proliferation and the extent time of post mitotic cell expansion determine the final size of leaf. Both of these growth phases are tightly regulated by a number of growth-promoting and growth-inhibiting factors

This figure is reproduced from Czesnick & Lenhard (2015) paper.

As the regulation of a leaf is affected by many factors, several defects can happen, that not necessarily produces a different phenotype. In flowering plants, there is a phenomenon called "*compensation*", which has been observed where a cell proliferation defect in developing leaf primordia triggers excessive cell expansion (Horiguchi & Tsukaya, 2011). The earliest study describing a compensation was by Haber in 1962 (Haber, 1962). In this study they describe a leaf development using gamma-irradiated wheat grains where the overall leaf development reduced and compensated with larger cells comparing to the control population of cells (Haber, 1962). Novel studies includes

similar findings of compensation for mutants that has a defect in positive regulators of cell proliferation, such as AINTEGUMENTA (Mizukami & Fischer, 2000), the manipulation of cell cycle proteins such as CDKA1 (Hemerly et al., 1995) and KRP2 (KIP-related protein 2) (De Veylder, 2001).

Other common phenotype of compensation is when smaller cell size appears in smaller organs, this called *more and smaller cells (msc)* phenotype first described in 2009 (Usami et al., 2009). The overall organ size is reduced, while the number of cells increased. According to the study, it seems that *msc* genes are not directly involved in the regulation of cell proliferation (Usami et al., 2009).

The third version of compensation when the compensating cell size can clearly hide the defective organ size. This is the case for *ant*- and *an3*-induced compensations (Horiguchi & Tsukaya, 2011). The overall cell number is significantly fewer, but cell size is larger compared to the wild-type cells (Horiguchi et al., 2005; Mizukami & Fischer, 2000). The actual result of these compensations is the increased proliferation (Mizukami and Fischer, 2000; Horiguchi et al., 2005). In these cases, the cause of compensation is clearly defective cell proliferation (Horiguchi & Tsukaya, 2011; Mizukami & Fischer, 2000). However it is interesting to see in *growth regulating factor 5* defective plants decrease cell number in the leaves, but fail to induce the same compensation despite it is a close interactor partner of *an3* (Horiguchi & Tsukaya, 2011; Narita et al., 2004).

As we can see, the cell size can compensate the organ or the plant size with increased proliferation. Because of these possible processes, I'm going to show that I investigated and measured the plant sizes in several dimensions not just the cell size.

4.1.3 Screening of cell size regulators

As I have reported in Section 2.3.2 earlier, that I have selected the Arabidopsis ortholog proteins of human TCTP, TP53RK, TNPO3, VPS18, CDC7 and the prefoldin group for screening of cell size regulators.

4.1.3.1 TCTP

4.1.3.1.1 In Plants

Translationally Controlled Tumour Protein (*TCTP*) is a known highly conserved, multifunctional protein that has been described to be involved in many fundamental biological processes and disorders either in human or other species (Bommer, 2017). Homologs of TCTP were identified in numerous plant species such as pea (Woo & Hawes, 1997), oil palm (Masura et al., 2011), rubber tree (Li et al., 2013) and rice (Wang et al., 2015) as well as in Arabidopsis (Berkowitz et al., 2008; Betsch et al., 2017; Schmid et al., 2005). In plants, TCTP is consist of 167-168 highly conserved sequence of amino acids, that are identical more than 70% in each plant organisms (Gutiérrez-Galeano et al., 2014).

4.1.3.1.2 In Arabidopsis

Arabidopsis homolog of TCTP protein, called AtTCTP, shows as high conserved amino acid similarity to human, Drosophila and yeast homolog of TCTP as 53.6%, 56%, and 62% (Betsch et al., 2017; Hinojosa-Moya et al., 2013; Thayanithy & Venugopal, 2005). TCTP protein showed to have a major role in the coordination of organ size and shape (Day & Lawrence, 2000). AtTCTP protein seems to be essential for the development of the final plant, while knock-out plants showed embryo lethality (Brioudes et al., 2010). TCTP mutant plants showed delayed development compared to wild-type plants (Brioudes et al., 2010). Silencing of the TCTP slowed down the vegetative development, while altered root development – shortened primary root length, and decreased lateral root formation by size and number (Berkowitz et al., 2008). Most crucially, TCTP mutants showed to have a decrease in cell size (Berkowitz

et al., 2008). TCTP seems to be involved in tumorigenesis in plants as well. However, it rarely manifests as physiological disorders (Doonan & Sablowski, 2010).

4.1.3.2 TP53RK (At1g12470) ortholog protein

At5g26110, the ortholog of human TP53RK protein, has not been fully characterised yet in Arabidopsis. It is currently known the TP53RK homolog protein in Arabidopsis holds an autophosphorylation activity (Nemoto et al., 2011). While it has also been described At5g26110 is one of the 72 calcium-dependent protein kinases that play a fundamental role in plant growth and plant development (Mittal et al., 2017).

4.1.3.3 Mos14

Mos14 (At5g62600), the ortholog of human TNPO3, was first characterised as a putative protein with similarity to transportin-SR protein (Bollman, 2003; Schlueter et al., 2003). Later the gene was described as a coding region of the Transportin-SR protein, a member of the importin-beta superfamily that imports serine-arginine rich proteins to the nucleus (Xu et al., 2011). MOS14 seems to be important in the proper slicing of genes that are important in plant immunity function (Xu et al., 2011). *Mos14* defective mutants showed to affect siRNA accumulation, transcriptional gene silencing and DNA methylation (Zhang et al., 2013).

4.1.3.4 VPS18 (At1g12470) ortholog protein

At1g12470, the ortholog protein of human VPS18 was identified in 2003 with the analysis of the vacuolar formation in Arabidopsis (Rojo et al., 2003). Functional analysis later confirmed, it contains a conserved RING domain such as its human counterpart (Stone et al., 2005). VPS18 ortholog in Arabidopsis is also a part of CORVET-specific, and HOPS-specific subunits, and plays a pivotal role in vacuolar cellular trafficking (Takemoto et al., 2018).

4.1.3.5 CDC7 (At4g16970) ortholog protein

At4g16970, the ortholog of human CDC7 protein, has been not fully characterised yet in Arabidopsis. The only published information is that the respective protein is a serine/threonine protein kinase categorised as Putative casein kinase II, with autophosphorylation activity (Nemoto et al., 2011).

4.1.3.6 The Prefoldin Group

Proteins in the prefoldin group have been identified in Arabidopsis in 2001 (Hill & Hemmingsen, 2001). Prefoldins are molecular co-chaperonins which are highly conserved from Archaea to Eukaryotes. (Martín-Benito et al., 2002; Rodríguez-Milla & Salinas, 2009). Prefoldins have reportedly coevolved with TCP-1 (also known as TCP-ring complex or TriC), actin and microtubules (Gu et al., 2008; Leroux & Hartl, 2000) as a name in yeast Genes Involved in Microtubule biogenesis (GIM) suggest (Rodríguez-Milla & Salinas, 2009). Prefoldins have two α subunits *Pfd3* and *Pfd5*, while four β subunits (*Pfd1*, *Pfd2*, *Pfd4* and *Pfd6*) (Siegert et al., 2000). Prefoldin proteins are present only in one copy in *A. thaliana*, while prefoldins show higher similarities to other prefoldins in other species than other subunits in Arabidopsis (Hill & Hemmingsen, 2001). Prefoldins seems to affect the orientation of microtubules (Locascio et al., 2013).

4.1.3.6.1 The function of alpha subunits (*Pfd3* and *Pfd5*)

Arabidopsis T-DNA mutants show that α subunits of the Prefoldin complex, *Pfd3* and *Pfd5* mutants, have significantly less α- and β-tubulins compared to the control Col-0 plants (Rodríguez-Milla & Salinas, 2009). It has been described that the location of *Pfd5* in Arabidopsis can be regulated by light-sensitive - phytohormones gibberellins via DELLA proteins (Dixit, 2013; Locascio et al., 2013). Gibberellins transmit signals to the nuclear-localised DELLA proteins, a protein family of nuclear growth-restraining proteins (Achard et al., 2007), then these proteins can hold or release *Pfd5* proteins (Dixit, 2013; Locascio et al., 2013). Nuclear localisation of PFD heterohexamer complex can limit microtubule stabilising effect and as a consequence can hinder plant growth and development (Locascio et al., 2013). *Pfd3* and *Pfd5* T-DNA mutants

showed alterations in the microtubule organisation and an essential role in salt stress tolerance (Rodríguez-Milla & Salinas, 2009). *Pfd3* and *Pfd5* mutants showed slightly darker colour, slower growth rate and smaller plants (Rodríguez-Milla & Salinas, 2009). Moreover, at cellular level *Pfd3* and *Pfd5* mutants showed alterations in the size and shape of pavement cells of the cotyledons (Rodríguez-Milla & Salinas, 2009). *Pfd3* and *Pfd5* mutants showed a slower growth rate of root development under experimental long day conditions (Rodríguez-Milla & Salinas, 2009).

Molecular evidence of *Pfd3*, *Pfd4* and *Pfd5* mutants show that PFDs negatively regulate HY5 transcription factor (Perea-Resa et al., 2017). HY5 a bZIP transcription factor that has a fundamental role in photomorphogenesis (Lau & Deng, 2010; Oyama et al., 1997), while it induces the biosynthesis of several genes that are needed for cold acclimation response (Schulz et al., 2015). Prefoldins interact with HY5 under cold conditions, while they are triggering their degradation via ubiquitination and attenuation of anthocyanin biosynthesis that results in the initiation of cold acclimation (Perea-Resa et al., 2017).

4.1.3.6.2 The function of beta subunits (*Pfd1*, *Pfd2*, *Pfd4* and *Pfd6*)

Pfd6 T-DNA mutants in Arabidopsis showed a series of visible defects and alterations (Gu et al., 2008). *Pfd6* mutants showed to have defects in cell division and cortical array organisation, while they show alteration in microtubule organisation and hypersensitivity to oryzalin (Gu et al., 2008). *Pfd6* mutants in Arabidopsis showed no interchangeable functions with other β subunits of the PFD complex such as *Pfd4* (Gu et al., 2008).

4.2 Materials and Methods

4.2.1 Plant material and growth conditions

Experiments with *Arabidopsis thaliana* plants were performed in the greenhouses of the research and innovation centre at the Fondazione Edmund Mach in San Michele all' Adige, Italy. Plants were grown and handled with the help of research assistants, and I performed all subsequent analyses by myself.

Arabidopsis experiments presented in this work were performed using wild-type and mutant plants in the Columbia (Col-0) genetic background. The GABI and SALK T-DNA insertion mutant lines listed in Table 4.2-1. were obtained from the Nottingham Arabidopsis Stock Centre (http://arabidopsis.info). Arabidopsis lines were either cultivated on GS90 soil (Manna Italia) or sterile half-strength Murashige and Skoog (½ MS) medium (Duchefa). Plants were grown at 22˚C under long day conditions consisting in 16^{h} of light at an intensity of 100 µmol m^{-2} s^{-1} light and 8^{h} of the dark.

4.2.2 Genotyping of Arabidopsis T-DNA insertion mutants.

Except for the EMS mutant *pfd6*, all T-DNA insertion homozygous mutant lines were obtained by screening the segregating T3 lines by PCR for the presence of the T-DNA insert using the primers indicated in Table 4.2-2. The PCR was performed using the 2 × PCR Super Master Mix (obtained from biotool.com) following the manufacturer's conditions on a Veriti thermal cycler (Applied Biosystems). The PCR cycle consisted on a denaturation step at 94˚C for 5 min followed by 35 cycles of 94˚C for 30 seconds, 55˚C for 30 seconds, 72˚C for 30 sec and a final elongation step of 5 min at 72˚C. The lines were considered homozygous if a PCR product was found using the primer combination T-DNA left border/genomic specific primer but not detected when using primers spanning the T-DNA insertion

Table 4.2-1: GABI and SALK T-DNA insertion mutant lines

Arabdidopsis locus id	Primer name in human	T-DNA insertion mutant lines
At4g16970	*cdc7*	SALK_130025, N683597
At2g07340	*pfd1*	GABI_689A09, N2047565
At3g22480	*pfd2*	SALK_041880C, N674646
At5g49510	*pfd3*	GABI_863G01, N340700
At5g49510	*pfd3/vbp1*	SALK_002489C, N681915
At1g08780	*pfd4*	GK345E03, N433075
At5g23290	*pfd5*	SALK_057848C, N671446
At1g29990	*pfd6*	N16396
At3g16640	*tctp1*	SALK_000005C, N670971 and N860169 (SAIL)
At5g62600	*tnpo3*	SALK_084257, N584257
At5g26110	*tp53rk*	SALK_071149, N571149 and N846003 (SAIL)
At1g12470	*vps18*	SALK_133060C, N654903

Table 4.2-2: Sequences of primers used in this work

Primer name	Sequence (5'-3')	Purpose
LBb1.3	ATTTTGCCGATTTCGGAAC	Left T-DNA border for SALK lines
GabiLBTDNA_o8474	ATAATAACGCTGCGGACATCTACATTTT	Left T-DNA border for GABI lines
F2g07340genoGabi	TGGTAAAATAATTTGGTGGTGTTAGC	Forward genomic primer on gene At2g07340. For genotyping pfd1 (GABI_689A09, N2047565) lines.
R2g07340genoGabi	TCTCCGAAGAATCATAGAGAGACC	Reverse genomic primer on gene At2g07340. For genotyping pfd1 (GABI_689A09, N2047565) lines.
Fgeno3g22480	ATGCAATTCGTGATTTCTTGG	Forward genomic primer on gene At3g22480. For genotyping pfd2 (SALK_041880C, N674646) lines.
Rgeno3g22480	ACGCTACCATAGACCGTGATG	Reverse genomic primer on gene At3g22480. For genotyping pfd2 (SALK_041880C, N674646) lines.
F5g49510genoGabi	TATTTTAAATCTCCGGGACCATAG	Forward genomic primer on gene At5g49510. For genotyping pfd3/vbp1 (GABI_863G01, N340700) lines.
R5g49510genoGabi	AGTAATACTTCAATAGCGGGGAAC	Reverse genomic primer on gene At5g49510. For genotyping pfd3/vbp1 (GABI_863G01, N340700) lines.
LP_SALK_002489	GAGTTCACTGAAGATCGCCTG	Forward genomic primer on gene At5g49510. For genotyping pfd3/vbp1 (SALK_002489C, N681915) lines.
RP_SALK_002489	CCGTAGGAGATCTGGGATTTG	Reverse genomic primer on gene At5g49510. For genotyping pfd3/vbp1 (SALK_002489C, N681915) lines.

F1g08780genoGABI	TATTTGGTGAATTCATTACGGGAT	Forward genomic primer on gene At1g08780. For genotyping pfd4 (GK345E03, N433075) lines.
R1g08780genoGABI	TTAATATAAGAAATCGCAGGCACG	Reverse genomic primer on gene At1g08780. For genotyping pfd4 (GK345E03, N433075) lines.
Fgeno5g23290	TGGTCCCAATTATTTTCTACCG	Forward genomic primer on gene At5g23290. For genotyping pfd5 (SALK_057848C, N671446) lines.
Rgeno5g23290	GTGTCCCAGGCACGTATAGAG	Reverse genomic primer on gene At5g23290. For genotyping pfd5 (SALK_057848C, N671446) lines.
LP_SALK_084257	GATTTTTCAGGTTCTCGAGGC	Forward genomic primer on gene At5g23290. For genotyping pfd5 (SALK_057848C, N671446) lines.
RP_ SALK_084257	CAAGCCAGAGACCAGTGAAAC	Reverse genomic primer on gene At5g23290. For genotyping pfd5 (SALK_057848C, N671446) lines.
LP_SALK_084257	GATTTTTCAGGTTCTCGAGGC	Forward genomic primer on gene At5g62600. For genotyping tnp53 (SALK_084257, N584257) lines.
RP_ SALK_084257	CAAGCCAGAGACCAGTGAAAC	Reverse genomic primer on gene At5g62600. For genotyping tnp53 (SALK_084257, N584257) lines.
LP_SALK_133060	CAGGCTTGACTGCTTCAGTTC	Forward genomic primer on gene At1g12470. For genotyping vps18 (SALK_133060C, N654903) lines.
RP_SALK_133060	AACTTCCTGAAGCCTTCAAGG	Reverse genomic primer on gene At1g12470. For genotyping vps18 (SALK_133060C, N654903) lines.

4.2.3 Reagents and equipment

Table 4.2-3. List laboratory reagents used in this work

Reagent name	Provider
1.5 ml microcentrifuge 'Eppendorf' tubes	StarLabs
15 ml and 50 ml Falcon tubes	VWR International Ltd
15 ml Pipettes	VWR International Ltd
Propidium Iodine (PI) at working concentration of 10 µg/ml	Invitrogen
PCR Super Master Mix	Biotool.com
Microscope slides	GE Healthcare

4.2.4 Analysis of plant development on soil

4.2.4.1 Leaf numbering

As Arabidopsis leaves are growing in a specific clockwise or anti-clockwise, leaf numbering was applied based on a protocol published by (Farmer et al. (2013). First two leaves, called cotyledons, that remain in the same size throughout in the adult life are not considered in the numbering. Leaf numbering is starting with leaf number 1 as shown in Figure 4.2-1.

Figure 4.2-1: Numbering of Arabidopsis leaves

An *Arabidopsis thaliana* plant is shown 21 days post germination (dpg). Leaves of the plant are coming in a specific order, clockwise or anti-clockwise (in this image clockwise). They can be numbered in order of appearance as it is stated.

Numbering is based on *Farmer et al. (2013).* The photo of the Arabidopsis plant was taken by me.

4.2.4.2 Measuring the plant size

Plants were grown on soil and photos were taken from above the growing plants 14 dpg every 2 days. Distorted photos have been corrected with Perspective Crop Tool from Adobe Photoshop CC software to make a same perspective analytical angle as indicated in Figure 4.2-2. Corrected angle photos were analysed using ImageJ software, Polygon selection tool. With this tool, triangles were drawn among the three longest leaves (leaf 7,8 and 9) then area and perimeter data were analysed. Three longest leaves are usually the same (leaf 7,8 and 9) and easily detectable. I used only those pictures of plants, which were having these three leaves clearly visible to reduce noise in my data.

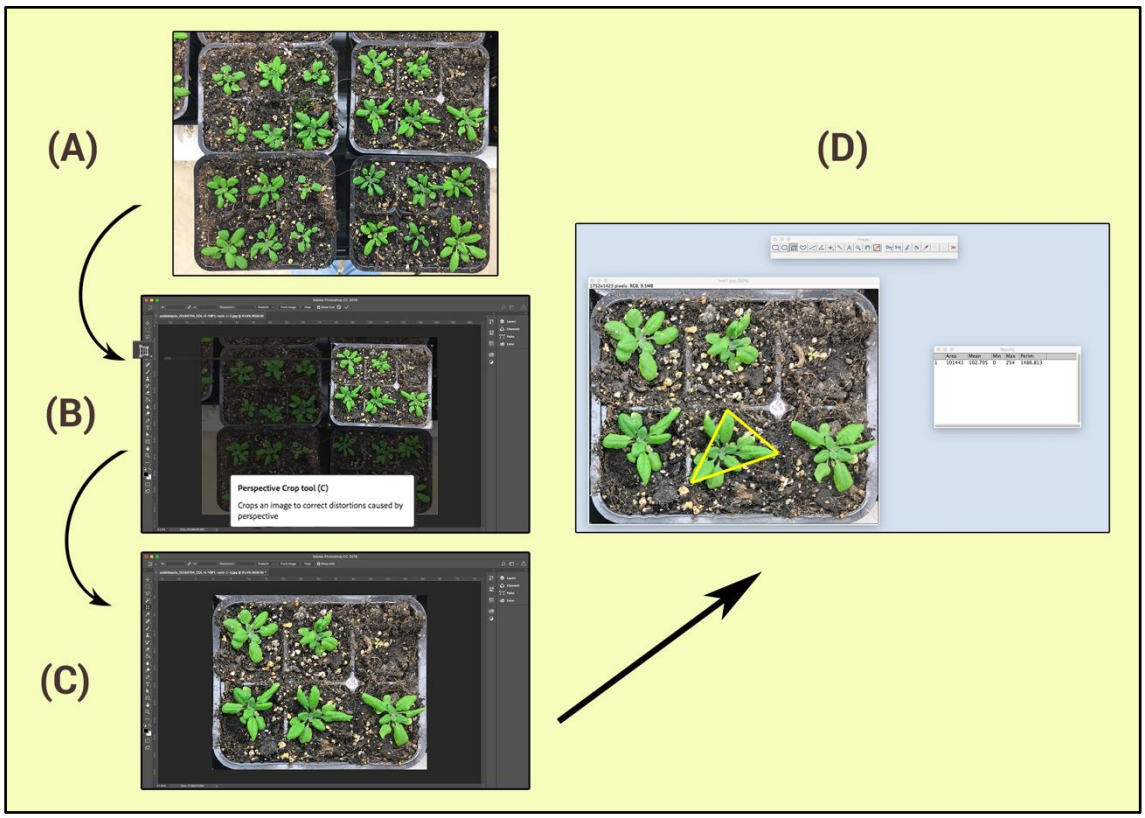


Figure 4.2-2: Measuring the plant size

Pictures that were taken with a different angle were corrected to have tall images normalised. **(A)** shows the original image. **(B)** Shows the cropping selection in Adobe Photoshop CC software, while **(C)** the results, fixed angle image. This image was later analysed using triangle methods in ImageJ software as shown in **(D)**.

4.2.4.3 Measuring the size of flower ramps

The main flower ramps of the Arabidopsis plants were measured 5 weeks / 35 days post germination (dpg). The length of the main flower ramp has measured with a manual ruler (Figure 4.2-3, A) and the number of later branches has been counted (Figure 4.2-3, B).

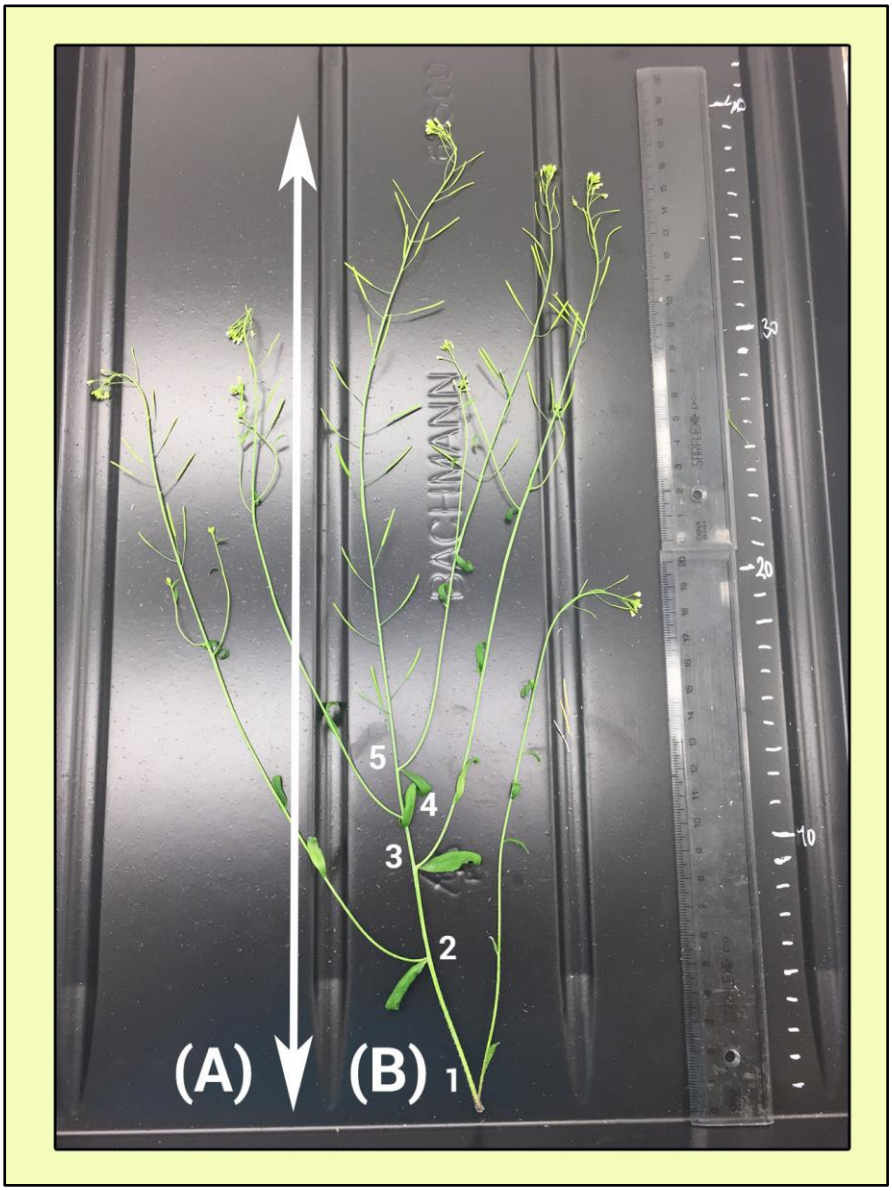


Figure 4.2-3: Measuring the development of flower ramp

(A) the length of the main flower ramp has been measured manually with a ruler.
(B) the number of later branches has also been counted.

4.2.4.4 Measuring the size of siliques and seeds

The size of siliques and seeds was measured four weeks post germination. For performing the measurement, siliques were taken from the main flower ramp between 10 and 20 cm from the base of the ramp (Figure 4.2-4, A). Pictures were taken using a dissecting microscope (Leica Microsystems MZ16 F stereo microscope), and photos were taken. Data was measured in ImageJ using straight line tool. Seeds were extracted from siliques from the main flower ramp between 10 and 12 cm and measured using two methods: (1) images were taken with Leica Microsystems MZ16 F stereo microscope, and ellipse shape superficies were measured in ImageJ with Oval Selection tool (Figure 4.2-4, B); (2) images were captured with a microscope using the Zeiss Axioscope and an AxioCam MR5 camera (Figure 4.2-4, C).. Images and measurement lines were taken with related AxioVision software, and the data later analysed in Microsoft Excel.

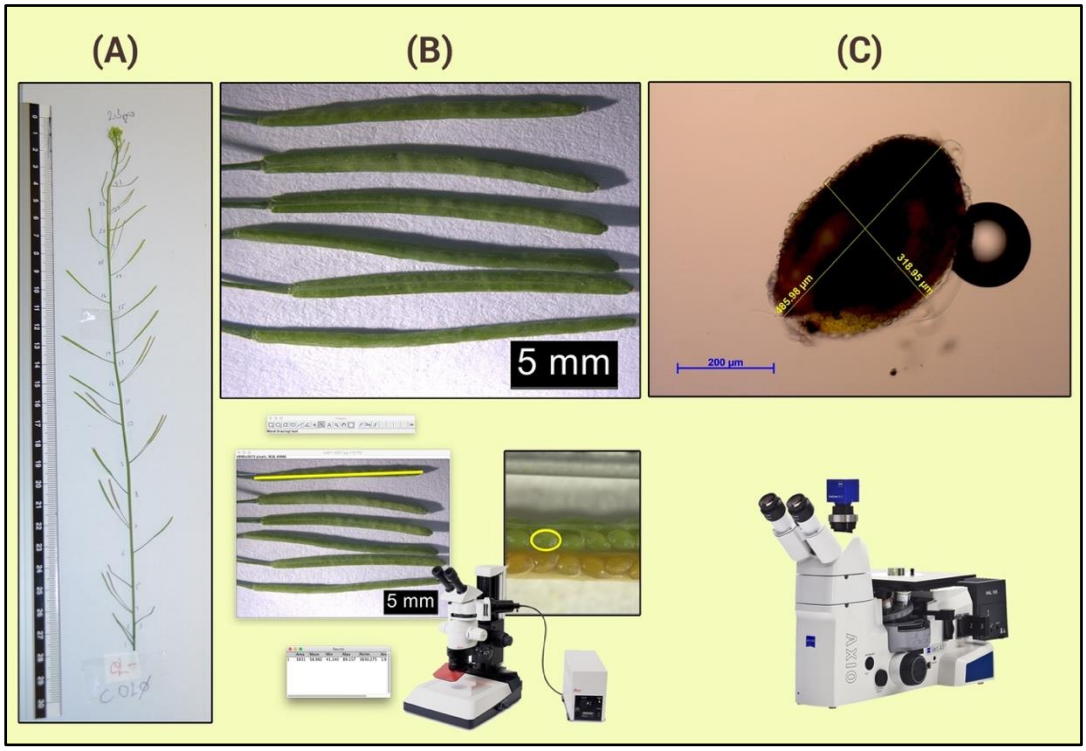


Figure 4.2-4: Measuring the size of siliques and seeds

(A) A main flower ramp from a plant at 28 dpg. Levels of siliques are numbered from the bottom to the top. **(B)** siliques between 10 and 20 cm from the base of the ramp and corresponding seeds between were measured with ImageJ software. **(C)** The size of seeds was also measured with *Zeiss Axioscope* and *AxioVision* software.

4.2.5 Analysis of plant development on sterile media

4.2.5.1 Analysis of leaf development

Photos from Arabidopsis plants grown in Petri dishes on sterile medium were taken at different time points after germination. Photos were taken from above the growing plants. Development of the cotyledons and the first leaf (#1) were analysed. Analysis was made using ImageJ software by applying a circle around the cotyledon or leaf area and the perimeter data calculated as shown in Figure 4.2-5.

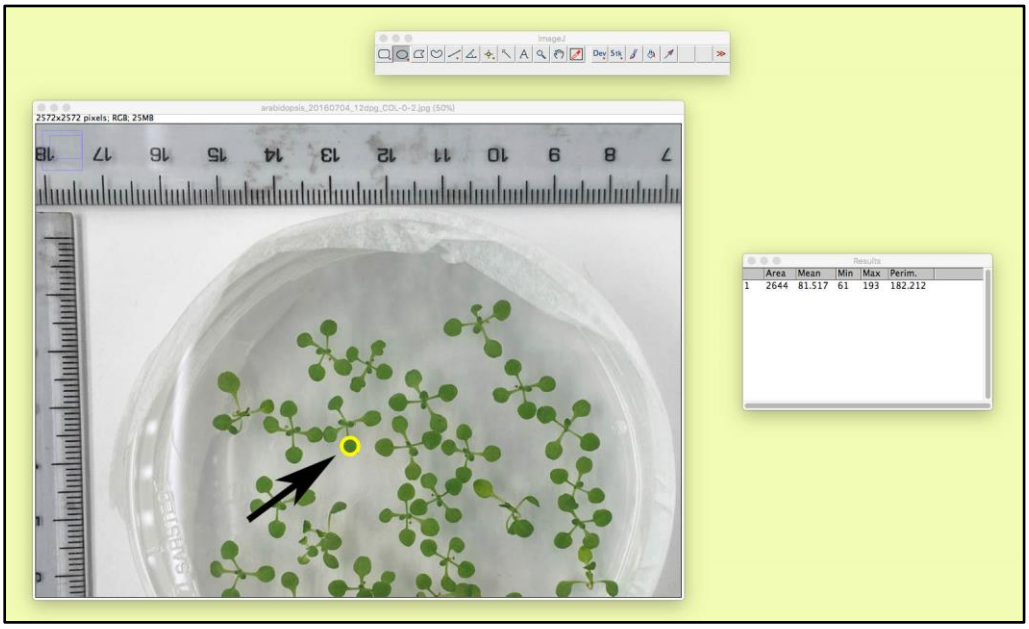


Figure 4.2-5: Analysis of leaf development in ImageJ software.

Black arrow pointing to a cotyledon that has been selected in ImageJ then circled with Oval Selection tool. The area was measured, and the perimeter calculated using Image J software.

4.2.5.2 Analysis of root development

Plates were positioned vertically, and the analyses were made on seedlings at different time points after germination every two days. I manually measured on each plate the length of the primary root (Figure 4.2-6, A) and number of lateral roots (Figure 4.2-6, B) using a ruler.

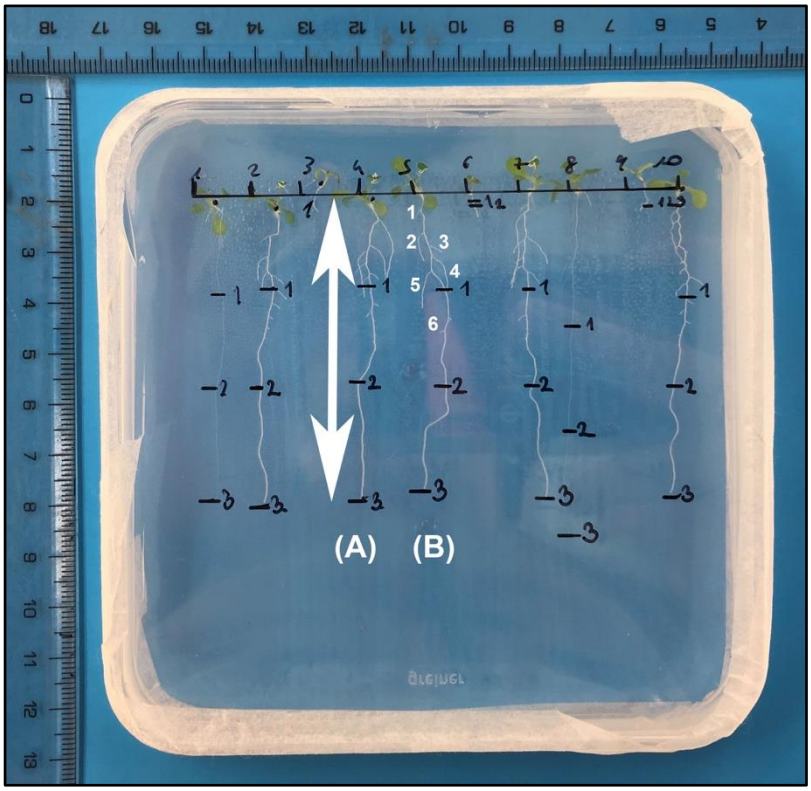


Figure 4.2-6: Development of roots

(A) Root length was measured manually using a general ruler and, while **(B)** the number of lateral roots has been counted.

4.2.5.3 Analysis of leaf epidermal cells under confocal microscopy

Seedlings were analysed 2 weeks / 14 days post germination (dpg). From each plant, the leaf number #5 was selected and cut. On the leaves specific areas under the half equatorial line were selected as shown in Figure 4.2-7, A. Leaf samples were incubated in propidium iodide (Invitrogen) for 2 minutes at 10 µg/mL in distilled water. Leaves were washed twice with water then put on microscope slides and analysed with a *Nikon A1* inverted confocal with a spectral detector (Figure 4.2-7, B). Images were made with Nikon's NIS Element software. Five images were taken at least from one leaf samples. In this specific part of the leaf, epidermal pavement cells grow into jigsaw shaped cells. Cell analyses were made with ImageJ software's Wand tool (Figure 4.2-7, C), measuring the perimeter and area data. Specific settings for threshold colour (red) were used.

Selection of leaves were always performed carefully as well as the cut for the jigsaw cells area. I was clearly paying attention to select only clearly visible cells and draw them precisely. In case the data is not correctly selected it can vary the result with high noise rate. I was doing all the selection of cells and all the analysis by myself to reduce the noise and create a consequent analysis.

4.2.6 Data Analysis

Recorded and measured data were analysed in Microsoft Excel. Basic statistical values were expressed as means, and their standard deviation and standard error were calculated. The levels of difference in standard deviation were calculated with F-test, then significance was calculated with two-tail Student's T-test.

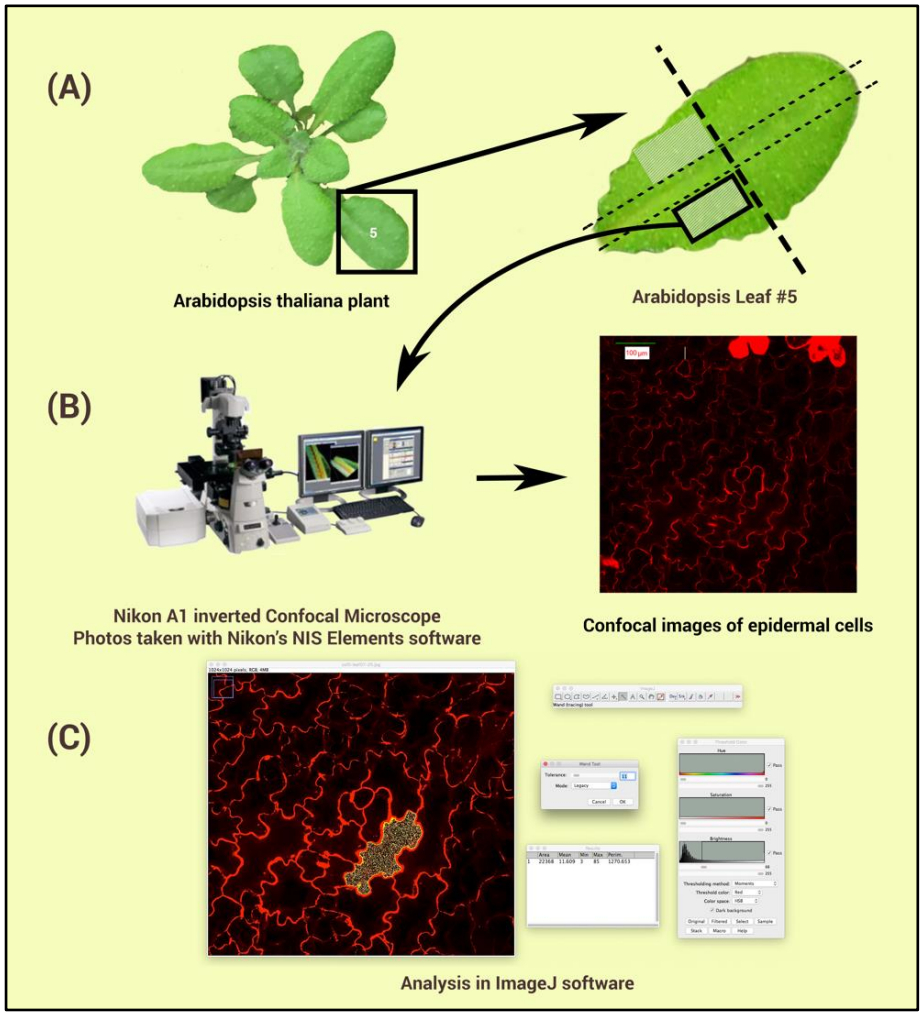


Figure 4.2-7: Analysis of Epidermal Cells

(A) Leaf number #5 was selected, and strips were cut diagonally **(B)** The leaf sectors were analysed, and pictures were taken using a Nikon A1 confocal inverted microscope. **(C)** Images were analysed with ImageJ's Wand tool, applied tolerance level between 10 and 15.

4.3 Results

4.3.1 Targets for screening

As stated in Chapter 2, I selected six different target genes plus the whole prefoldin group proteins in Arabidopsis to evaluate using a reverse genetic approach using T-DNA insertion mutant lines and assess their role in cell size regulation in the plant. These are poorly studied proteins that have not been reported to regulate cell size in *Arabidopsis thaliana* yet. Proteins are listed in Table 4.3-1. Three of the genes have not been named yet and therefore used the name of the corresponding human ortholog for the Arabidopsis gene (*cdc7*, *tp53rk* and *vps18*). *Vps18* gene has been selected but not tested systematically. As well as the *pfd4* KO mutant has been selected, but not tested, while a homozygous were available in the two separate seed batches. I tested *cdc7*, *mos14*, *pfd3*, *tctp1* and *tp53rk* mutants, plus the whole prefoldin complex excluding prefoldin 4: *pfd1*, *pfd2*, *pfd3*, *pfd5* and *pfd6*.

The knock-out seeds of these proteins were ordered from Nottingham Arabidopsis Stock Centre (NASC) and tested by growing the plants on the soil. Only checked homozygous knock-out (KO) mutants were tested. For a comparative growth see Figure 4.3-1.

In the upcoming experiments shown below, Col-0 was used as a wild type, as well as a negative control. I used Col-0 for all of my experiments to see any change. In the prefoldin experiments *pfd6* mutants were used as a positive control for size.

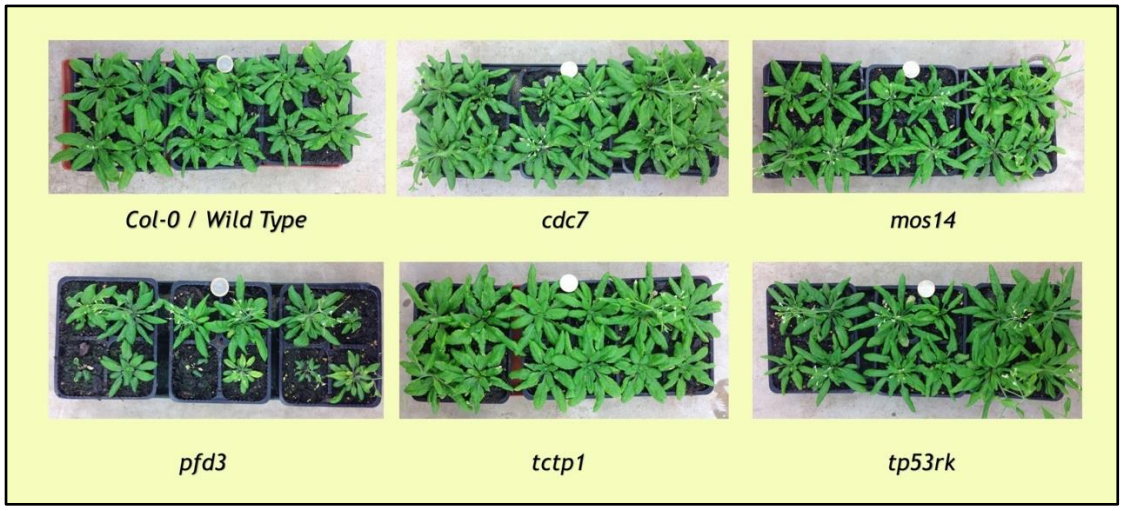


Figure 4.3-1: Phenotypic characterisation of the Arabidopsis T-DNA mutant lines

Cdc7, *mos14*, *pfd3*, *tctp1* and *tp53rk* mutants and the wild-type Col-0 grown on soil were tested. Pictures were taken 28 days post germination (dpg). During this screening *pfd3* mutant showed a distinct phenotype (smaller plants) compared Col-0 control plants, but we were unable to observe this phenotype in later experiments. This screening shows just an early comparison of different knock-out phenotypes.

Table 4.3-1: List of proteins for the screenings

Arabidopsis Protein Name	Arabidopsis Locus Name	Description from TAIR database	Human ortholog's Protein Name
n/a	At1g12470	*Zinc ion binding protein*	VPS18
PFD1	At2g07340	*Prefoldin Subunit 1*	PFD1
PFD2	At3g22480	*Prefoldin Subunit 2*	PFD2
PFD3	At5g49510	*Prefoldin Subunit 3*	VBP1
PFD4	At1g08780	*Prefoldin Subunit 4*	PFD4
PFD5	At5g23290	*Prefoldin Subunit 5*	PFD5
PFD6	At1g29990	*Prefoldin Subunit 6*	PFD6
n/a	At4g16970	*Protein kinase superfamily protein*	CDC7
MOS14	At5g62600	*Encodes a nuclear importer of serine-arginine rich (SR) proteins and is involved in the regulation of splicing of R genes by regulating the import of the SR proteins into the nucleus.*	TNPO3
n/a	At5g26110	*Protein kinase superfamily protein*	TP53RK
TCTP1	At3g16640	*Encodes a protein homologous to translationally controlled tumour protein (TCTP) from Drosophila.*	TCTP

4.3.2 Analysis of Plant Development

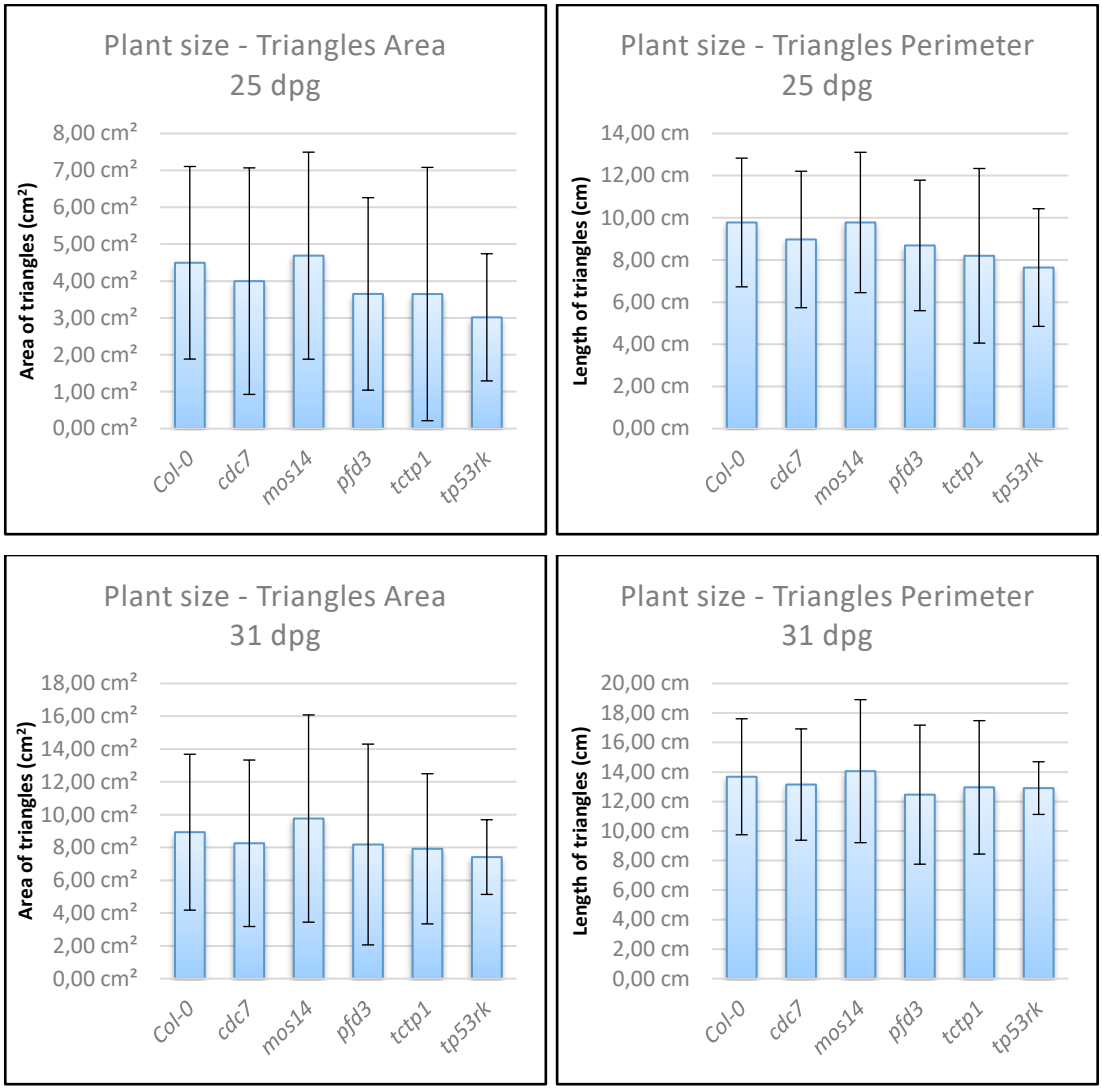
I performed a systematic analysis of *cdc7*, *mos14*, *pfd3*, *tctp1* and *tp53rk* T-DNA insertion KO mutant lines as well as the prefoldin group (*pfd1*, *pfd2*, *pfd3*, *pfd5* and *pfd6*) TDNA KO mutant lines.

Here in this section I show the analysis of plant, leaf and root sizes.

4.3.2.1 Plant size

First, I analysed and compared the plant size of the mutants to one of the wild-type Col-0 plants at 25 dpg and 31 dpg. I found that there is no significant difference in plant size area or perimeter at these particulate time points as indicated in Figure 4.3-2.

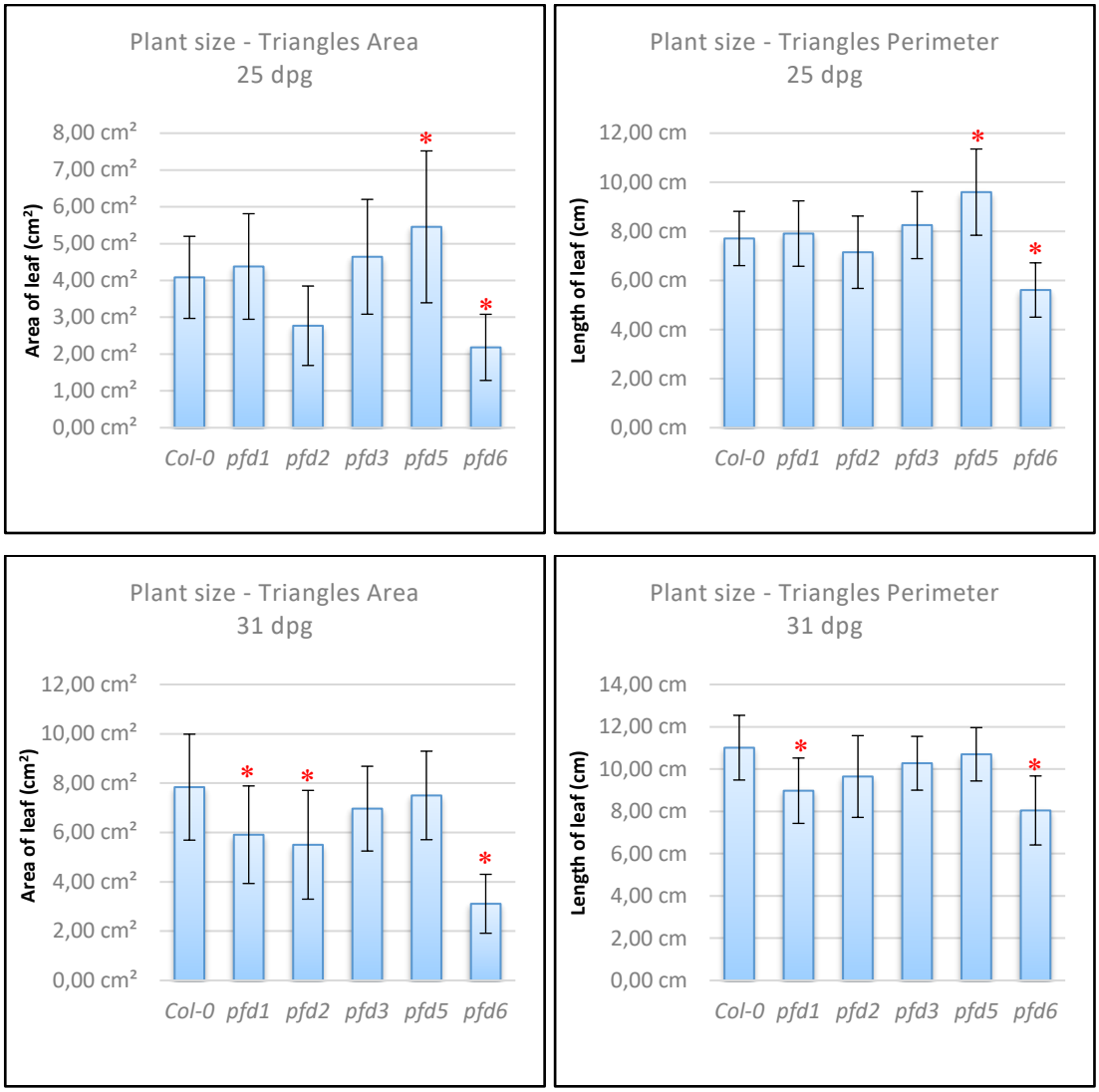
Moreover, I measured the area and perimeter of plants of the prefoldin complex. I found that *pfd6* - the negative control of my experiments - is significantly smaller in all experiments. On another hand, *pfd5* KO mutant was larger at 25 dpg, and *pfd1* KO mutant was smaller at 31 dpg. (Figure 3.3-3). *Pfd3* KO mutant showed consistently no difference to wild-type population. These results were different, comparing to *Pfd3* and *Pfd5* mutants that have been reported to show consistently smaller plants (Rodríguez-Milla & Salinas, 2009).



Sample Nrs	25 dpg	31 dpg
Col-0	11	11
cdc7	8	9
mos14	9	11
pfd3	11	9
tctp1	11	7
tp53rk	8	8

Figure 4.3-2: Measuring the plant size

Plants were grown on soil, and perimeter values were calculated at 25 and 31 dpg using the triangle measurement method. **There were no significant differences**. This experiment was repeated two individual times with no significant difference between the size of the mutants and the corresponding wild-type Col-0.



Sample Nrs	25 dpg	31 dpg
Col-0	34	37
pfd1	40	44
pfd2	17	16
pfd3	41	39
pfd5	32	26
pfd6	36	37

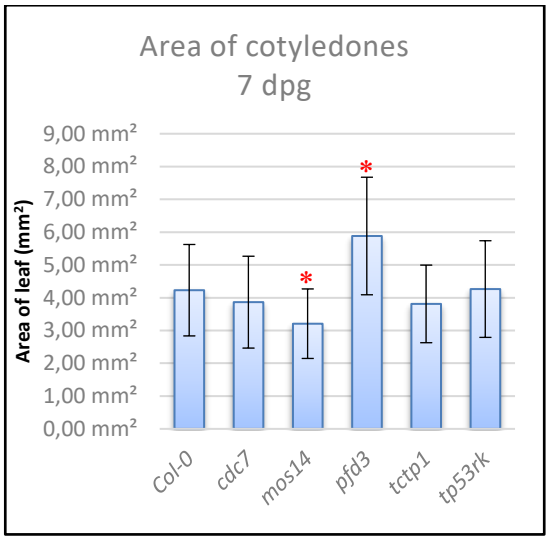
Figure 4.3-3: Plant size of prefoldins

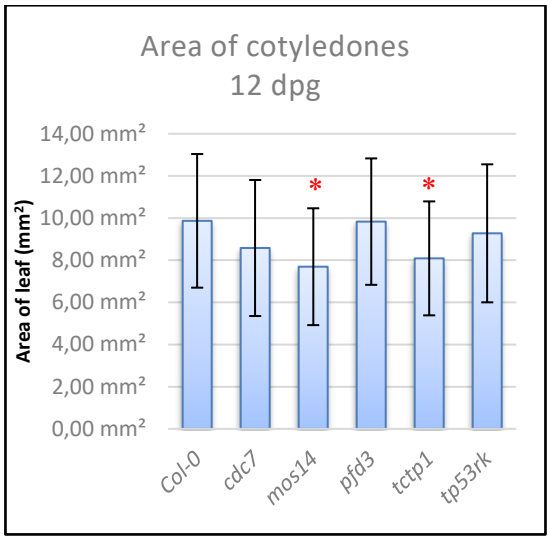
Plants were grown on soil; the size was calculated using the triangle method at 25 and 31 dpg. The p*fd6* KO mutants were significantly smaller in all parameters tested while *pfd5* KO mutants were larger (25 dpg) and *pfd1* KO mutants were smaller (31 dpg). The experiment was repeated two individual times with consistent results.

* denotes significance level <0.05

4.3.2.2 Leaf size - Cotyledones

I analysed and compared the sizes of cotyledons 7 and 12 dpg (Figure 4.3-4) for *cdc7*, *mos14*, *pofd3, tctp1* and *tp53rk*. I found that *mos14* KO plants are significantly smaller and that *pfd3* KO plants have significantly larger leaves than the wild-type Col-0 at 7 dpg. 5 days later at 12 dpg, I found that *mos14* KO and *tctp1* KO plants have significantly smaller leaves.





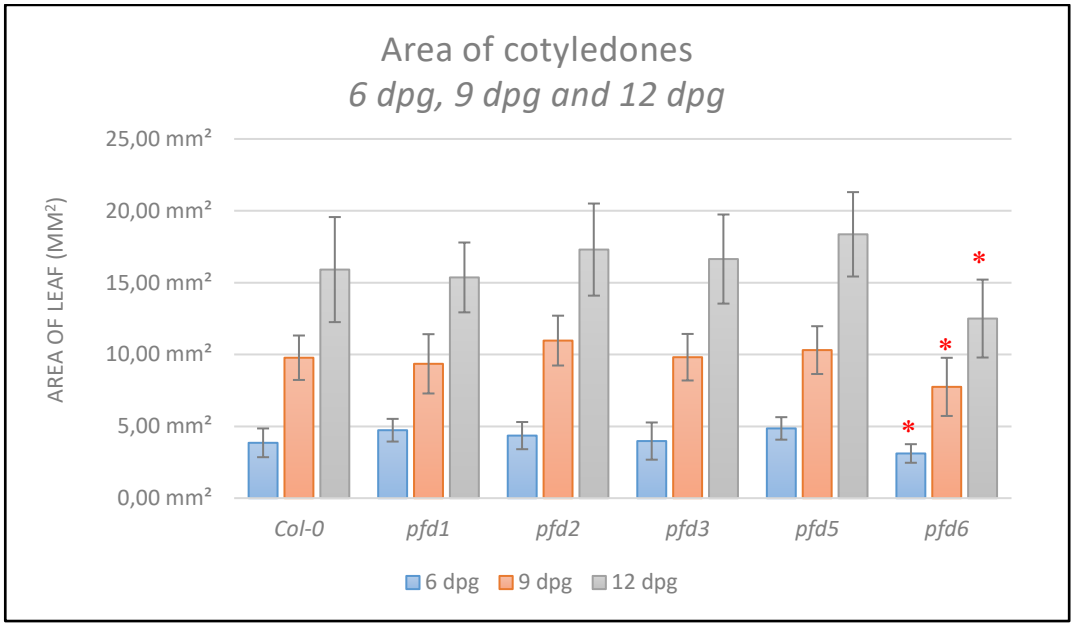
Sample Nrs	7 dpg	12 dpg
Col-0	31	27
cdc7	22	29
mos14	29	27
pfd3	36	35
tctp1	30	31
tp53rk	27	28

Figure 4.3-4: Comparing the size of cotyledons

The area of the cotyledons was measured 7 dpg (left) and 12 dpg (right). The *mos14* KO mutants are significantly smaller (12 dpg), while *pfd3* KO mutants are significantly larger at12 dpg. The *mos14* and *tctp1* KO mutants are significantly smaller (12 dpg). The experiments were repeated two times Independently with consistent results.

* denotes significance level <0.05

Then, I analysed the size of cotyledons at 6,9 and 12 dpg for *pfd1*, *pfd2*, *pfd3*, *pfd5* and *pfd6*. I found that cotyledons of *pfd6* KO mutants – the negative control of my experiments, are significantly smaller in every experiment (Figure 4.3-5), while leaves of *pfd3* KO mutants showed significantly larger leaves in half of the experiments (not shown on Figure 4.3-5). However, none of the experiments showed significantly larger cotyledons for *Pfd5* mutants as it has been reported (Rodríguez-Milla & Salinas, 2009). In the experiment below, the cotyledons of *pfd5* looks a little bit larger, but it is not significant.



Sample Nrs	6 dpg	9 dpg	12 dpg
Col-0	34	34	0
pfd1	26	34	32
pfd2	28	33	32
pfd3	26	34	32
pfd5	30	34	32
pfd6	30	33	30

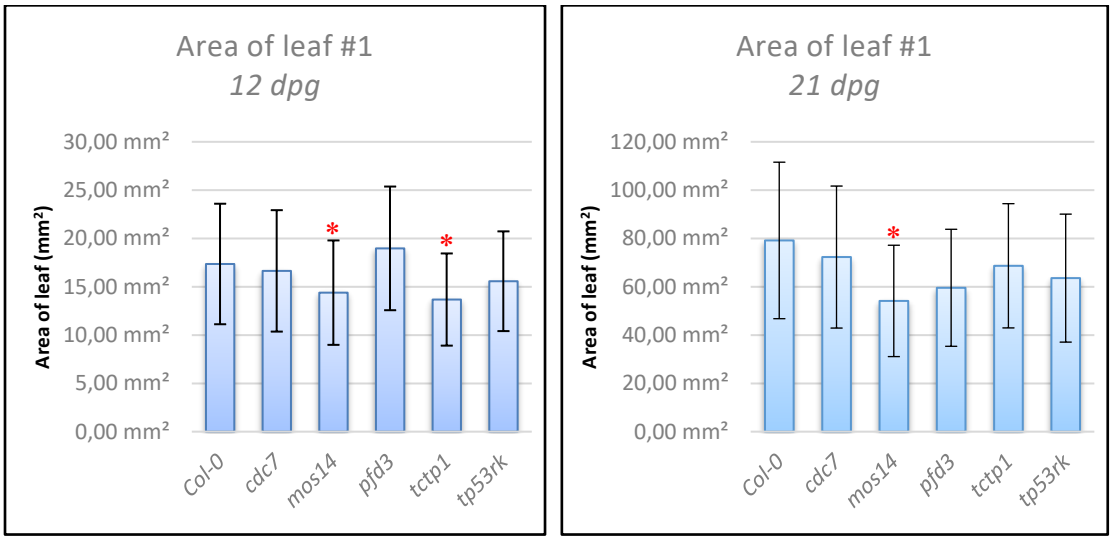
Figure 4.3-5: Cotyledon size of prefoldin mutants

Cotyledon leaf size was checked at 9 and 12 dpg. The negative control, *pfd6* KO mutants were consistently smaller during the 12 days of growth.

* denotes significance level <0.05.

4.3.2.3 Leaf #1

I analysed and compared sizes of leaf number 1 at 12 dpg and 21 dpg for *cdc7*, *mos14*, *pfd3*, *tctp1* and *tp53rk* (Figure 4.3-6). I found that *tctp1* and *mos14* KO plants have significantly smaller leaves at 12 dpg and 9 days later at 21 dpg, I found that *mos14* KO plants have still significantly smaller leaves.



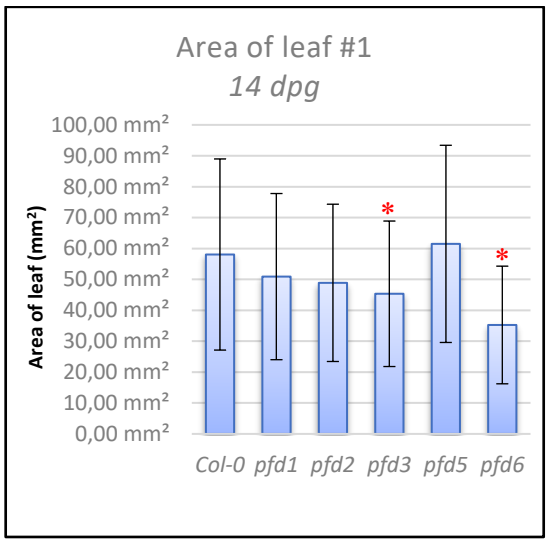
Sample Nrs	12 dpg	21 dpg
Col-0	24	16
cdc7	23	18
mos14	21	15
pfd3	31	16
tctp1	25	20
tp53rk	26	15

Figure 4.3-6: Size of leaf #1

Area of leaf number #1 was measured at 12 (left) and 21 dpg (right). The *tctp1* and *mos14* KO mutants are significantly smaller at 12 dpg. The experiments were repeated independently two times with consistent results.

* denotes significance level < 0.05

Leaf #1 was measured and counted 14 dpg for *pfd1*, *pfd2*, *pfd3*, *pfd5* and *pfd6.* I found that *pfd3* and *pfd6* KO mutants are significantly smaller in leaf size comparing to the Col-0 plants (Figure 4.3-7). I did not follow up this experiment 9 days later.



Sample Nrs	14 dpg
Col-0	28
pfd1	31
pfd2	28
pfd3	30
pfd5	27
pfd6	28

Figure 4.3-7: Size of leaf #1 of prefoldins

The size of leaf #1 was checked at 14 dpg. The *pfd3* KO mutants were significantly smaller than the wild-type Col-0.

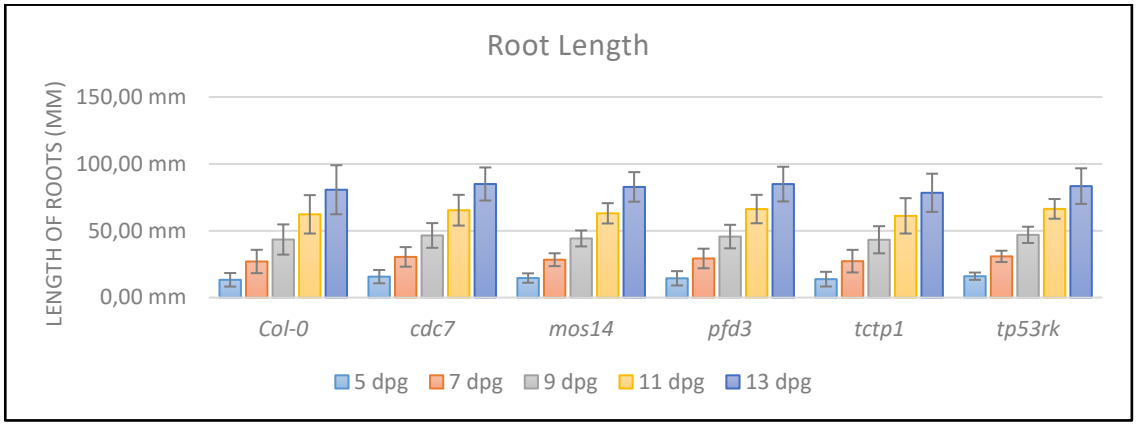
* denotes significance level <0.05.

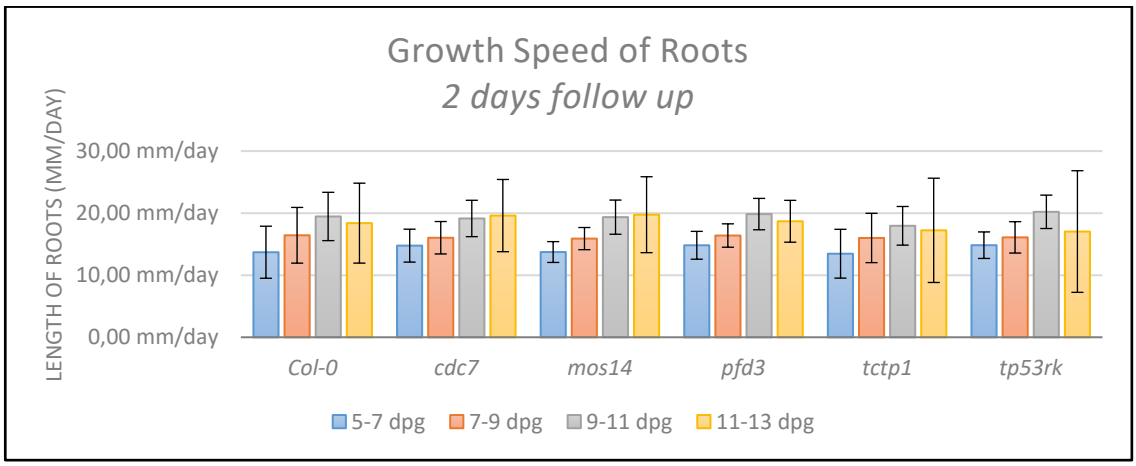
4.3.2.4 Root sizes

I analysed and compared the root size changes at 5, 7, 9, 11 and 13 dpg for *cdc7*, *mos14*, *pfd3*, *tctp1* and *tp53rk*. I found that there was no significant difference in the length, the rate of root growth and number of lateral roots (Figure 4.3-8). However, in later experiments, *pfd3* KO mutant showed shorter roots at a slower speed rate (Figure 4.3-8).

Morover, I analysed the length of root, the growth rate and the number of lateral roots for *pfd1*, *pfd2*, *pfd3*, *pfd5* and *pfd6* (Figure 4.3-9). Unlike in earlier analyses, I found that *pfd3* KO mutants are significantly shorter and grew at a lower speed. The *pfd6* KO mutants were significantly shorter and slower as well. The *pfd1* KO mutants showed some significant differences in length at some stages (day 5), but this was not consistently observed.

These results are different from what was described in the literature since my experiments did not show any alterations or lagging in size and growth rate for *Pfd3* and *Pfd5* T-DNA mutants (Rodríguez-Milla & Salinas, 2009).





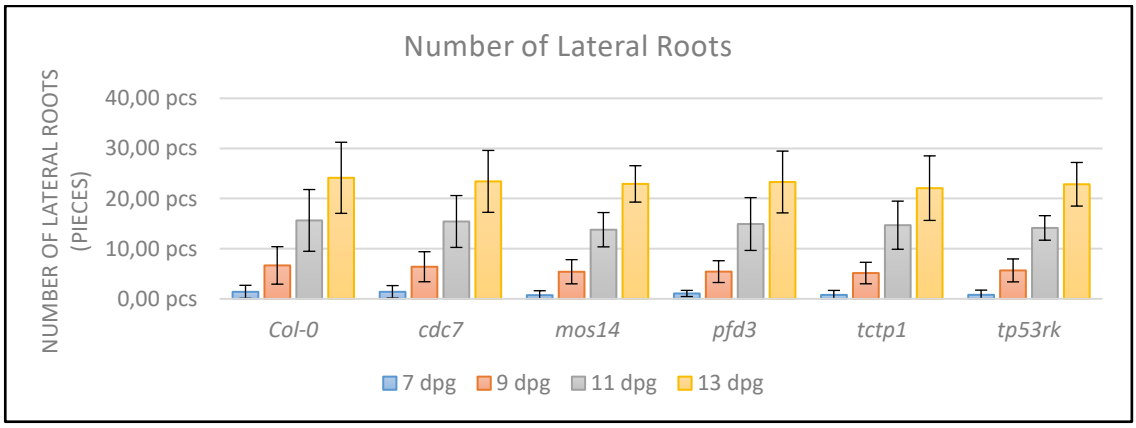
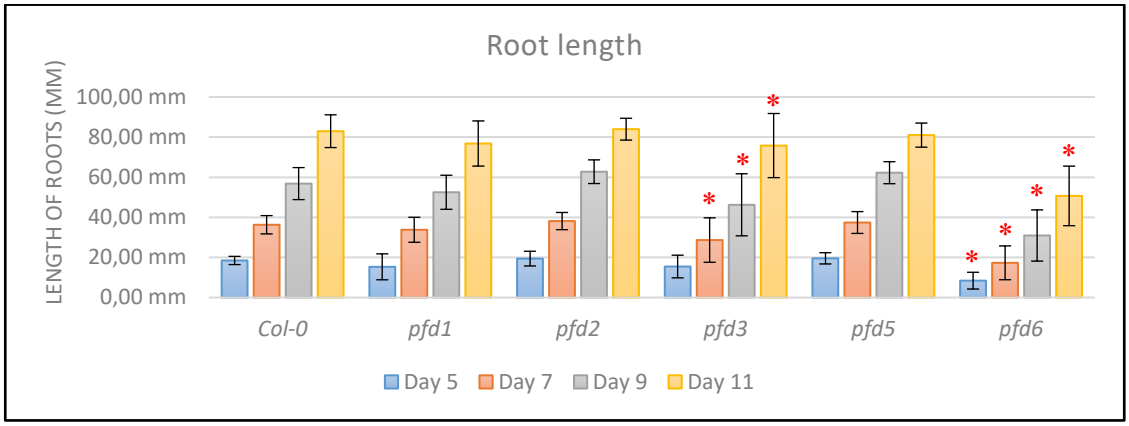
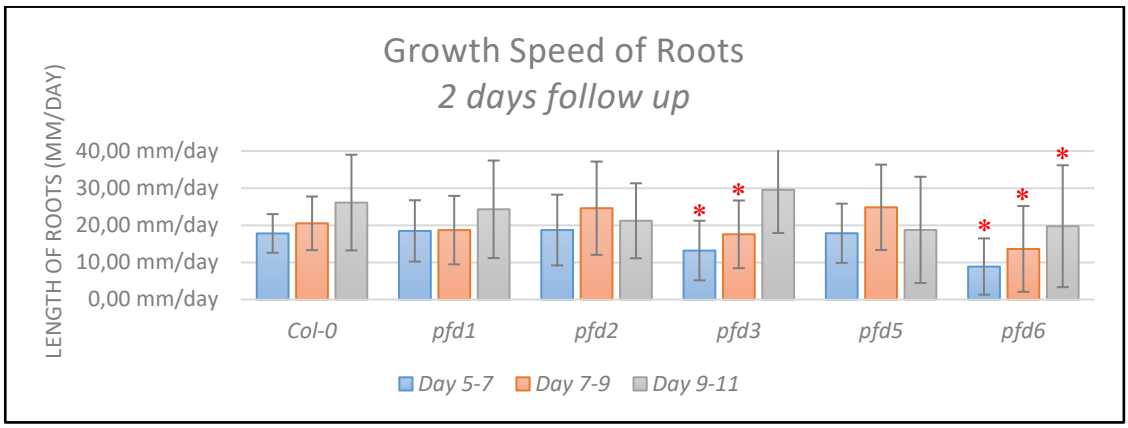


Figure 4.3-8: Measuring the root size

Twenty individual mutant plants were grown next to each other in a petri dish. Sample numbers were between 12 and 20 for each data point. Root length (top chart), the speed of growth (middle chart) and a number of lateral roots (bottom chart) were measured. **There was no significant difference** between any mutant comparing to the *Col-0* samples in this experiment.





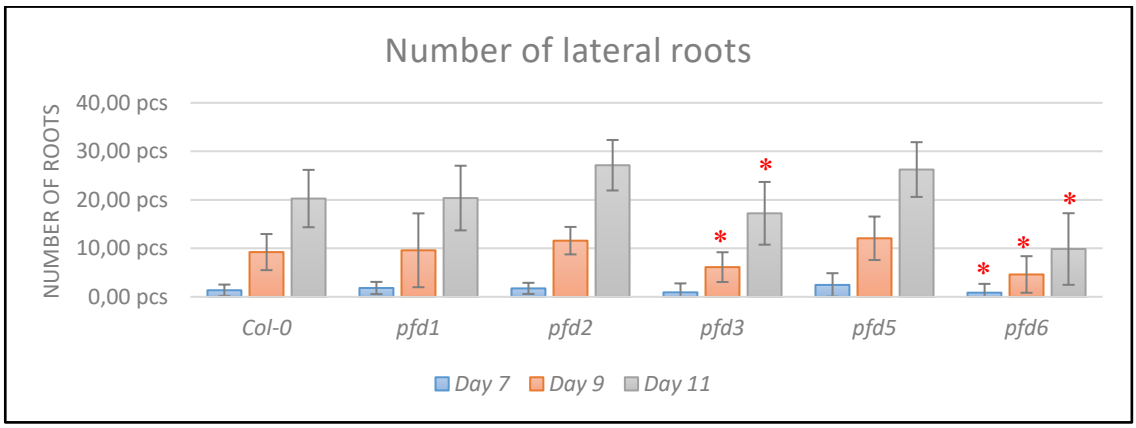


Figure 4.3-9: Root length, growth rate and number of lateral roots in prefoldin mutants

The pfd3 and *pfd6* KO mutants were significantly shorter and slower in growth.

* denotes significance level <0.05.

4.3.2.5 Measuring the flower ramp

I analysed the main flower ramps for *pfd1*, *pfd2*, *pfd3*, *pfd5* and *pfd6*. I found that *pfd1*, *pfd3* and *pfd6* KO mutants have significantly (<0.05) shorter main flower ramps then Col-0. See Figure 4.3-10 for the details.

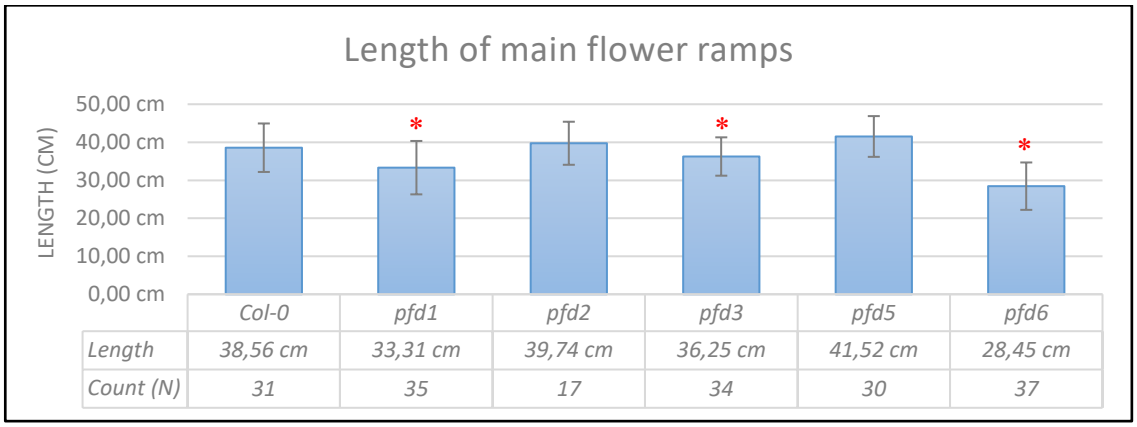


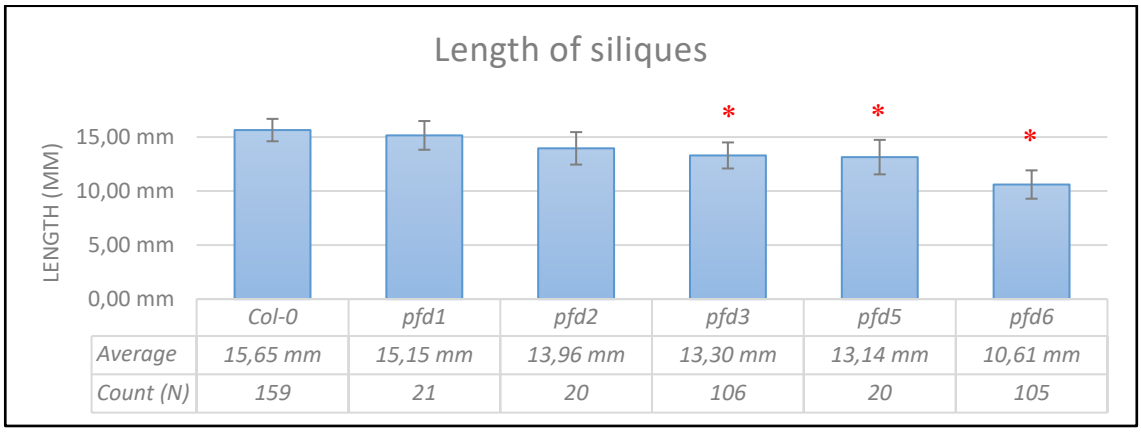
Figure 4.3-10: Length of main flower ramps

Pfd1, *pfd3* and *pfd6* KO mutants were significantly shorter flower ramps 42 dpg.

* denotes significance level <0.05.

4.3.2.6 Silique size

I analysed the size of the siliques in the prefoldin mutants as well and found that all of the prefoldins have shorter silique, although only *pfd3*, *pfd5* and *pfd6* KO mutants have significantly shorter siliques (taken from the main flower ramp, between silique 10 and 12 cm from the base of the ramp as shown in Figure 4.3-10.



	Col-0	pfd1	pfd2	pfd3	pfd5	pfd6
Average	15,65 mm	15,15 mm	13,96 mm	13,30 mm	13,14 mm	10,61 mm
Count (N)	159	21	20	106	20	105

Figure 4.3-11: Length of siliques in prefoldin mutants

Pfd3, *pfd5* and *pfd6* KO mutants have significantly shorter siliques.

* denotes significance level <0.05.

4.3.2.6.1 *pfd3* mutant and silique size

As *pfd3* mutant showed significantly smaller or shorter phenotype in some dimension, I analysed and compared the size of the siliques. Plants were grown next to each other. From each plant, five siliques were harvested at the same distance 10 cm above the base of flower ramp as explained in material and methods. Interestingly, *pfd3* KO mutants showed a clear phenotype distinct from the wild-type and the other mutant lines. In Figure 4.3-12, the results of 20 different individual experiments along with the summary statistics are shown. It can be clearly seen that the outcome results of individual experiments are highly variable, despite the plants were kept at the same light, temperature and watering conditions. The only reason behind could be the differences on soil material however I autoclaved the soil before the use.

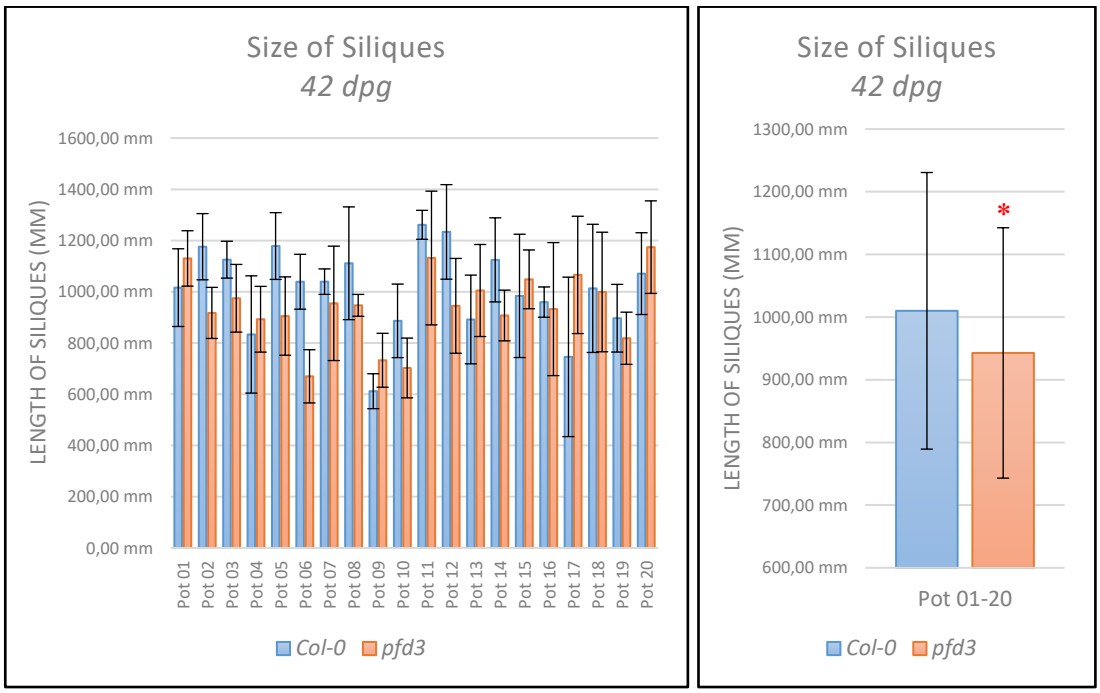


Figure 4.3-12: Comparing the size of siliques

The results of 20 individual experiments are shown the right panel and summarised on the left. The number of samples was 100 siliques for both Col-0 and *pfd3* KO mutants. The p*fd3* KO mutants are significantly shorter.

* denotes significance level < 0.05

I found that the length of siliques on *pfd3* KO mutants are significantly shorter ($p < 0.05$) compared to the Col-0 plants a side to side comparison for three individual plant pots from the same levels of siliques at the main flower ramp above 10 cm is shown in Figure 4.3-13.

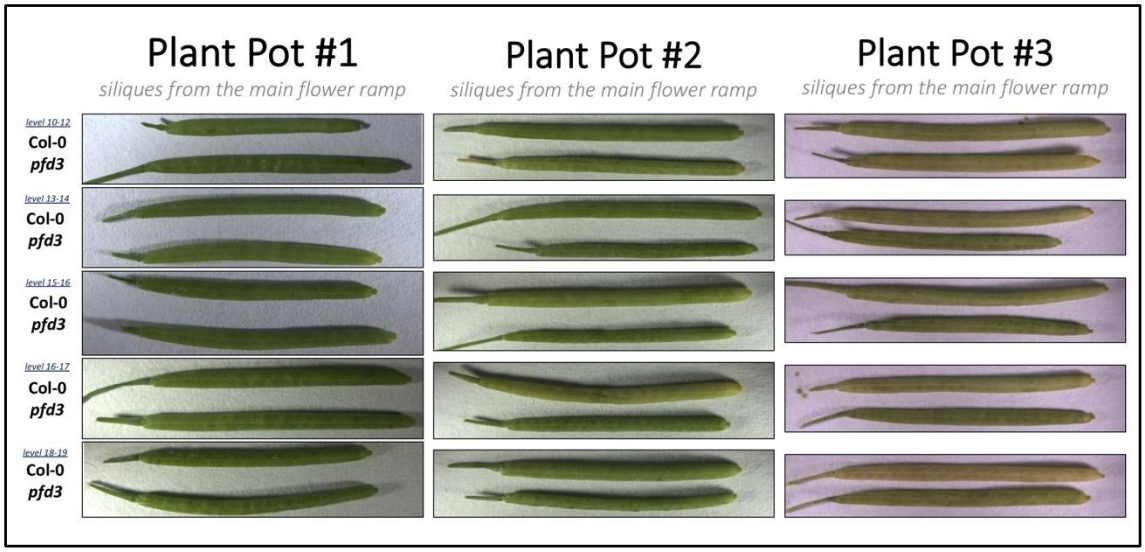


Figure 4.3-13: Comparison of siliques

Siliques were harvested from the same level of the main flower ramp in Col-0 and *pfd3* KO mutants. Mutants were grown next to Col-0 in the pot. The figure shows three individual pots for visual comparison.

The *pfd3* KO mutant showed an intriguing phenotype that can be explained as a "*bumping*" that was not observed in Col-0 plants. This phenomenon did not occur in the other KO mutants (Figure 4.3-14).

We named this phenotype as "*bumping*" with my colleagues, while I did not find any relevant term in the current literature. I mean under this term that the seeds are clearly visible from the unopened silique, because of the larger seed size in the silique.

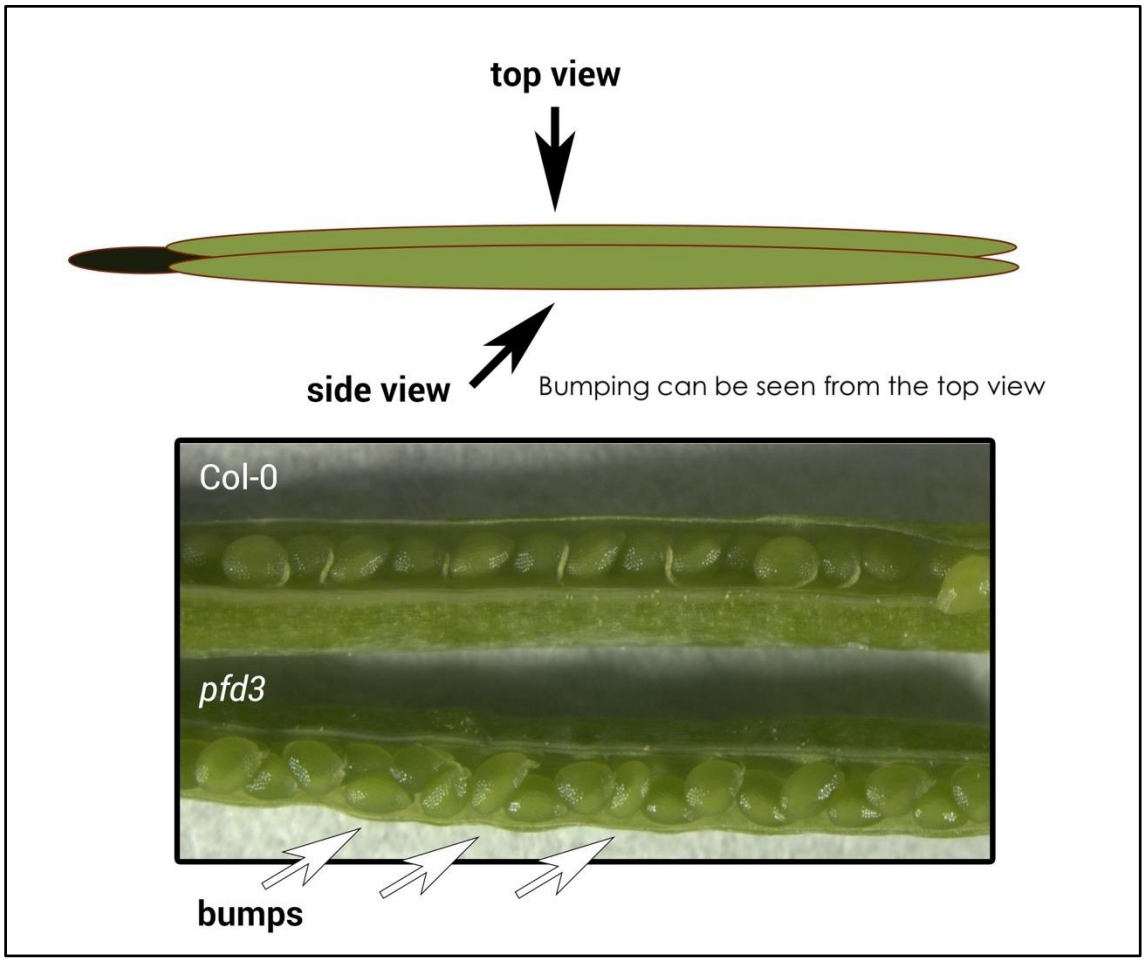


Figure 4.3-14: The "bumping" silique phenotype of pfd3

Bumps on siliques of the *pfd3* KO mutant compared to the wild-type *Col-0*.

4.3.3 Seed size

4.3.3.1 *Pfd3* mutant and seed size

The phenotype was reproducible in many different independent experiments and was further investigated (Figure 4.3-15, top part). I measured individual seed sizes for the KO mutants and the control plants. I found that *pfd3* KO mutant seeds are 8.85 % larger if 2D oval shape measuring was performed using ImageJ (Figure 4.3-15, bottom part).

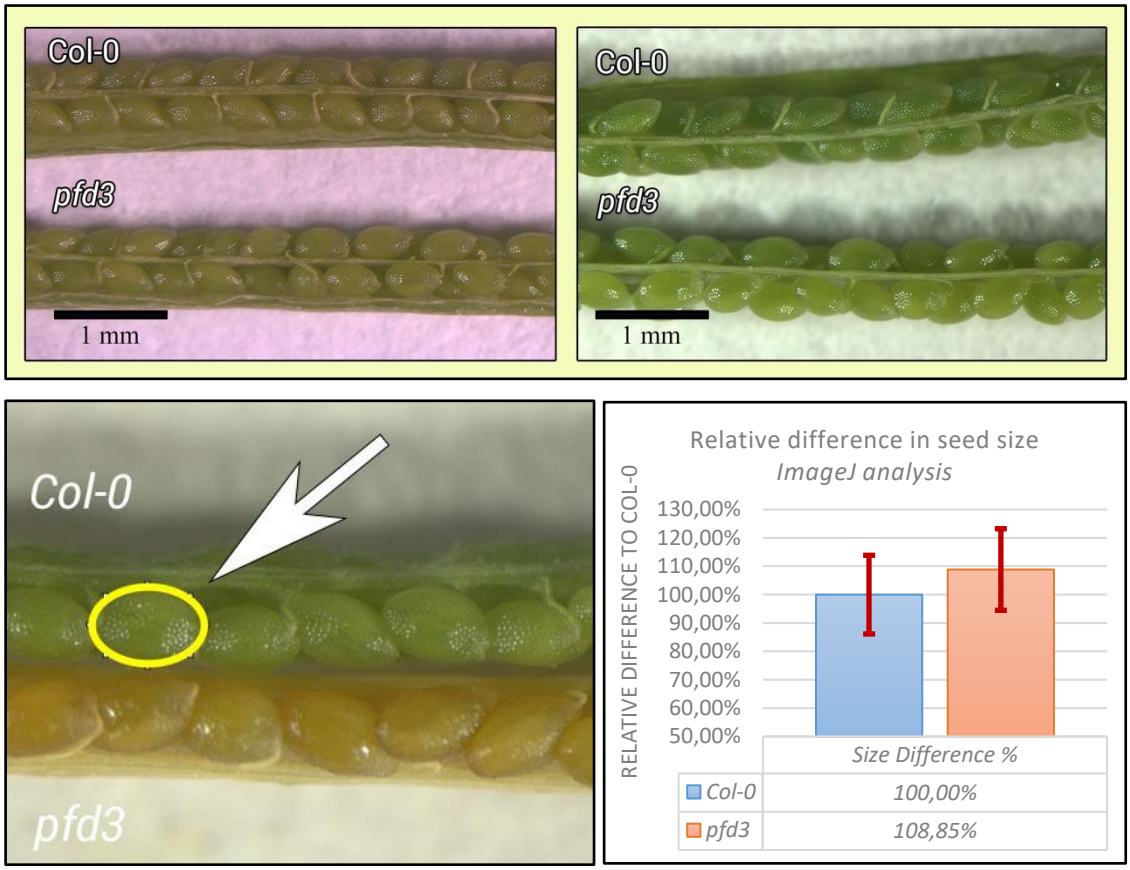


Figure 4.3-15: Measuring the size of seeds with ImageJ

Pfd3 KO mutant seeds were larger in 2D oval shape measuring with 8.85% and 13.5% in theoretical volume. The calculation based on 141 seed sample measurements from 10 individual plants for both *pfd3* KO mutant and *Col-0*.

Microscopic analysis of *pfd3* KO mutants later confirmed these findings. The *pfd3* are larger by 7.99% in 2D surface and 10.76% in 3D volume as shown in Figure 4.3-16.

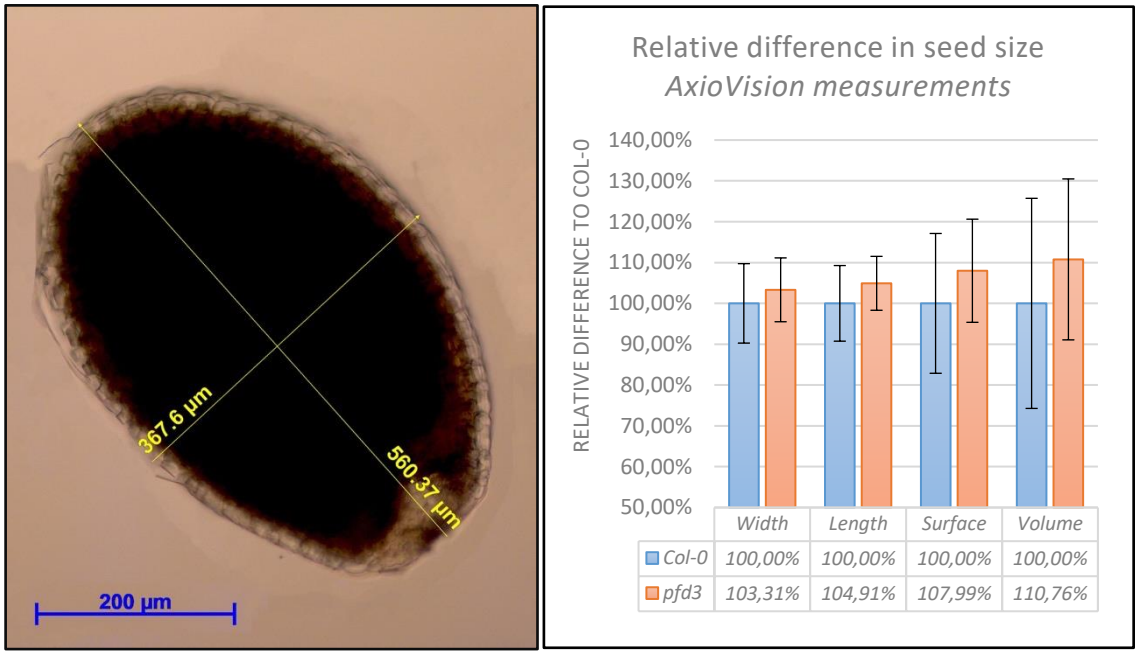


Figure 4.3-16: Microscopical measurements of seed size

Pfd3 KO mutant seeds were larger in width (3.31%) and length (4.91%). As well as the derivative 2D surface with 7.99% and in 3D volume with 10.76%. The calculation based on 141 seed sample measurements from 7 individual plants for both *pfd3* KO mutant and *Col-0*.

4.3.3.2 *Pfd6* and seed size

I checked the cell size of prefoldins mutants and found that in addition to the *pfd3* KO mutants showed earlier, *pfd6* KO mutants have significantly smaller dimensions in seed size as indicated in Figure 1.1-17. As I used *pfd6* as a negative control during my experiments I was aware of the notion that these mutants should have a smaller phenotype in leaf and root sizes, however the smaller seed size was a novel result.

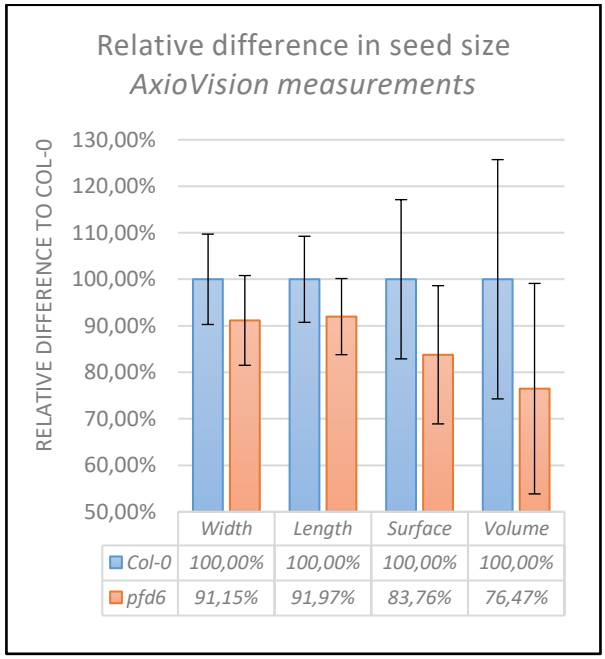


Figure 4.3-17: Seed size in *pfd6* KO mutants

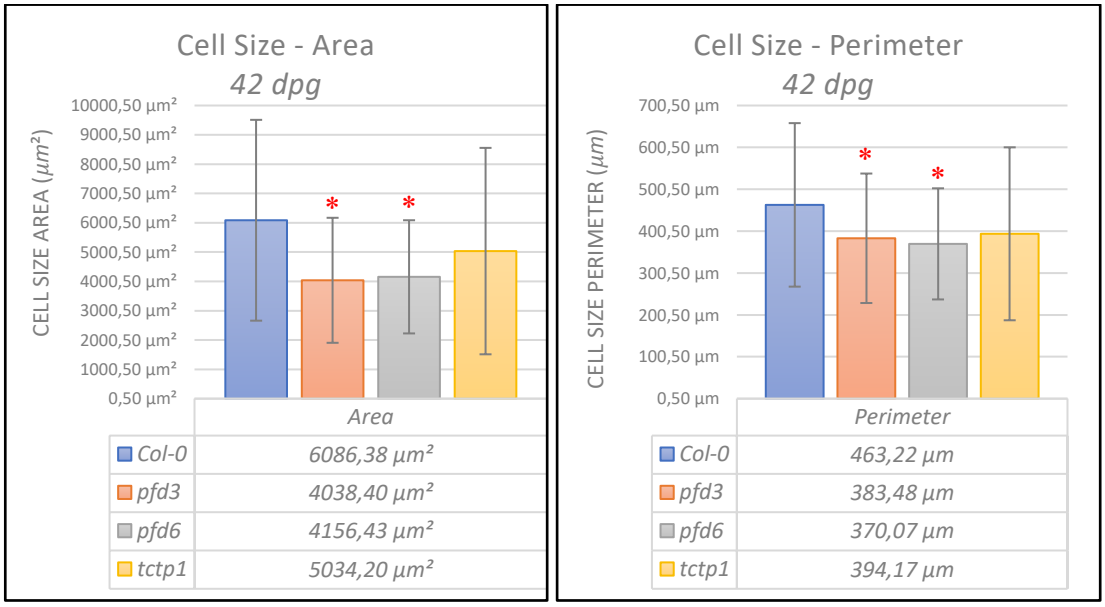
Seed size in *pfd6* KO mutants is significantly smaller. These plants are 8.85% smaller in width, and 8.03% smaller in length compare to Col-0. This means the 2D surface is smaller about 16.24% percent and the theoretical volume is 23.53% compared to Col-0 plants.

4.3.3.3 Other genes

Tctp1 and *tp53rk* KO mutants were checked and analysed for seed size, but there was no significant difference in seed size comparing to the Col-0 plants.

4.3.4 Measuring the cell size: *pfd3*, *pfd6* and *tctp1*

While I have analysed nearly every developmental aspect of *Arabidopsis'* organs in all prefoldin mutants, I completed the phenotypical characterisation by assessing if the KO mutation would affect the pavement epidermal cell size. These cells show jigsaw-like shapes in Arabidopsis. I analysed *pfd3*, *pfd6* and *tctp1* KO mutants' cell size by measuring the area and perimeter data of epidermal cells of plants at growth saturation *i.e.* when starting flowering. The size of the epidermal cells was calculated from confocal microscopy pictures. I found that *pfd3* and *pfd6* KO mutants have significantly smaller cell size, while *tctp1* KO mutants have about the same cell size than Col-0, however, with high variability (Figure 4.3-18).



	Col-0	*pfd3*	*pfd6*	*tctp1*
Area	6086.38 µm²	4038.40 µm²	4156.43 µm²	5034.20 µm²
St. Diff. Area	3425.27 µm²	2132.37 µm²	1930.57 µm²	3520.86 µm²
Perimeter	463.22 µm	383.48 µm	370.07 µm	394.17 µm
St. Diff. Perimeter	195.21 µm	154.60 µm	132.67 µm	206.51 µm
Sample Nr.	110	152	98	206
Leaf Nr.	9	8	8	9

Figure 4.3-18: Analysis of Cell size

Leaves were harvested 6 weeks / 42 days post germination. I found that size of the cells in *pfd3* and *pfd6* KO mutants are significantly smaller.

Note for this experiment: I had time for one complete experiment only, while I got promising and significantly different data, I asked the help of my Italian supervisor **Dr Azeddine Si-Ammour**, who repeated this experiment for me. Here I showed data of his experiment. He grew the plants, he took the images, and I did only the analysis of cell size, using ImageJ software, where I got very similar results.

In my experiment, I measured **4669.79 µm²** (± 2528.57 µm²) for *pfd3* KO mutant epidermal cells, from 954 cell size samples (15 leaves), while here I calculated **4038.40 µm²** (± 2132.37 µm²) based on 152 samples from 7 leaf samples. Based on these experiments I can say that *pfd3* KO mutants are significantly smaller than wildtype, Col-0 plants.

4.4 Discussion

In this chapter, I showed the results obtained using A*rabidopsis thaliana* and regarding cell size and growth measurements using T-DNA insertion mutants for the genes *cdc7*, *mos14*, *pfd3*, *tctp1*, *tp53rk* and the six gene members the prefoldin family. On one hand, I obtained interesting results for some mutants like *pfd3*. However, on the other hand, I showed that the phenotype associated with some mutants contradicts what was described in the literature.

4.4.1 Prefoldin group

The Prefoldin complex contains six proteins from *pfd1* to *pfd6*. The complex is known to fold newly synthesised polypeptide chains, including actins and tubulins (Geissler et al., 1998; Vainberg et al., 1998) and play a fundamental role in the maintenance of cellular homeostasis (Martín-Benito et al., 2007; Siegert et al., 2000). In *Arabidopsis*, I tested the effects of *pfd1*, *pfd2*, *pfd3*, *pfd5* and *pfd6* knock-out T-DNA mutants. *Pfd4* gene was not tested and not shown, while I was not able to obtain homozygous plants for the mutation.

4.4.1.1 *Pfd3* and *Pfd5* – Causing changes in plant and cell size

As reported earlier, the prefoldin *pfd3* and *pfd5* mutants had alterations in the size and shape of cotyledon pavement cells (Rodríguez-Milla & Salinas, 2009). The growth rate of root development of these mutants was also slower under the same experimental conditions used in my study (Rodríguez-Milla & Salinas, 2009). I reported above that I found larger cotyledon leaves for the *pfd3* mutant, as described by Rodríguez-Milla *et al.* (2009). However, I did not observe a slower rate of growth or alterations in root development and overall size changes (Rodríguez-Milla & Salinas, 2009). In contrary, I found that *pfd3* mutants show significantly smaller cells, the opposite that has been reported by Rodríguez-Milla & Salinas (2009). Neither did I observe a significantly different phenotype for the *pfd5* mutant from that reported by Rodríguez-Milla &

Salinas (2009). Further experiments need to be done using different growth conditions to determine whether there are technical reasons for the differences.

4.4.1.2 *Pfd3* – Causes changes in seed size

In addition to the phenotype described above for the prefoldin *pfd5* mutant, I found that the *pfd3* T-DNA mutant has significantly larger seed sizes. This phenotype has never been reported before. The seed size is an essential attribute for reproduction (Sundaresan, 2005). A seed consists of three distinct components: the embryo, the endosperm and a coat that covers the seed (Gasser et al., 1998; Grossniklaus et al., 1998). Production of seeds is coordinated in a regulated manner, and it is triggered by fertilization (Alonso-Blanco et al., 1999; Gasser et al., 1998; Grossniklaus et al., 1998). Different methods have been used to look for genes that regulate seed size in plants. First is to discover these by a general screening or by quantitative trait loci (QTL) (Herridge et al., 2011). For example, Alonso-Blanco *et al.* (1999) identified six QTL that has an effect on seed size without significant effects on the plant (Alonso-Blanco et al., 1999). Seed size can be regulated *via* integument elongation by regulating ARF2 (Auxin Response Factor 2) (Schruff, 2005) and TTG2 (Transparent Testa Galbra 2) (Johnson, 2002), or *via* control of early endosperm proliferation, involving MINI3 (Miniseed3) (Luo et al., 2005) and Haiku 1-2 (Garcia, 2003). Knocking out genes encoding these four examples of seed size regulators caused a decrease in the size of the seeds. My results show that *pfd3* T-DNA mutant has significantly larger seeds. This phenotype has been reported previously for *da1* gene overexpression (Li et al., 2008). The *da1* gene encodes a ubiquitin receptor and is thought to limit the final size of the seeds as well as plant organs (Li et al., 2008). A similar change in seed phenotype has been recently reported using the loss-of-function mutant of *raptor1b*, which functions in the TOR pathway (Salem et al., 2017). Producing larger seeds is potentially important for food products and further work is needed to understand the mechanisms that control seed size, particularly of agriculturally important plants.

4.4.1.3 *Pfd6* – Control for small plants

Pfd6 is one of the four β-subunits of the prefoldin complex (Martín-Benito et al., 2007). It has been reported by Gu *et al.* (2008) that *pfd6* mutants of *Arabidopsis* have a series of microtubule defects that affect cell division, cortical array organization and microtubule dynamics (Gu et al., 2008). They also showed that knockout plant have reduced levels of tubulins (Gu et al., 2008). In my experiments, *pfd6* mutants were used as a positive control plant. I was able to reproduce all the reported details of *pfd6* mutants, such as defects in cell division and alteration in microtubule organization that causes visible phenotypes (Gu et al., 2008). I showed that *pfd6* mutants were significantly smaller, grew at a slower speed, have shorter roots and shorter main flower ramps. As *pfd6* mutants are still viable, it raises the question whether *pfd6* is not essential for growth or that the mutant can be compensated, perhaps by other prefoldin proteins (Gu et al., 2008). Another possibility could be the regulation of beta-tubulins *via* the chaperonin TCP1 (CCT), that also regulates microtubule formation in a prefoldin-independent manner (Castellano & Sablowski, 2008).

4.4.2 *Tctp1* – A mutant with known size control

The experiments using the T-DNA insertion mutant *tctp1* (At3g16640) did not result in any relevant phenotypes, despite the fact that all mutants I used were homozygous (as checked by PCR). These homozygous *tctp1* mutants were the same as the wild-type population in every developmental aspect, whether of the roots or the leaves. Previously, *tctp1* knock-out plants were retrieved *via* an embryo-rescue approach and showed a severe delay in development and critical growth defects with small organs and short plant stature (Brioudes et al., 2010). These mutants also flowered late and were sterile (Brioudes et al., 2010). Berkowitz *et al.* (2008) also described *tctp1* mutants with an impaired vegetative devolvement and a leaf expansion that was slowed down due to reduced cell size. However, I run at least 3 complete independent experiments on different SALK mutants; I was unable to identify the embryo-lethal *tctp1* T-DNA insertion mutant phenotype described by others in the literature (Berkowitz et al., 2008).

I focused my further study on leaf measurements, whereas Berkowitz *at al.* (2008) proposed that *tctp1* mutation is embryo-lethal due to a defect in the male gametophyte (Berkowitz et al., 2008). Later, Brioudes *et al.* (2010) later showed a conflicting result stating that fertilization still occurs in embryo-rescue knock-out plants (Brioudes et al., 2010). Recently Hafidh *et al.* (2016) repeated this experiment, and they described pollens of wild-type and *tctp1* mutants that are not as different as described previously, while fertilization between the *tctp1* mutant and wild-type still occur (Betsch et al., 2017). Furthermore, Brioudes *et al.* (2010) showed that the TCTP protein in *Arabidopsis* controls cell proliferation (Brioudes et al., 2010). These findings were confirmed by other research groups working on tomato, tobacco and cabbage (Betsch et al., 2017; Bruckner et al., 2017; Cao et al., 2010; Gupta et al., 2013; Tao et al., 2015). The reasons behind the difference in phenotype observed between the *tctp1* mutants used in all these studies and mine could be the nature of mutation used to knock-out the *tctp1* or other technical differences in the conditions used during the experiment's gene. Further work is necessary to understand and characterise this discrepancy.

4.4.3 *Mos14* – A mutant with smaller leaves

MOS14 protein is an important receptor for nuclear-import of serine and arginine-rich splicing factors (Xu et al., 2011). *Mos14* mutants have been described to cause defects in siRNA accumulation and transcriptional silencing (Zhang et al., 2013). I found that the size of the cotyledons and leaf #1 are significantly smaller (< 0.05) in *mos14* T-DNA mutants compared to the Col-0 population. This is the first study reporting such leaf and cotyledon size phenotypes. The human ortholog, *TNPO3*, has been reported to act as a similar serine/arginine-rich splicing factor importer to the nucleus, including the alternative splicing factor 1/pre-mRNA-splicing factor SF2 and polyadenylation-specific factor (CPSF6) (Lai et al., 2001; Lee et al., 2010). The reason behind the size difference in leaves could be the change in nuclear transport, as has been reported for the HASTY protein (Bollman, 2003). HASTY is also a transporter protein, and loss-of-function mutation causes reduction of the size of roots and lateral organs in *Arabidopsis* (Bollman, 2003).

4.4.4 Arabidopsis orthologs of CDC7 and TP53RK

These genes are poorly characterized in *Arabidopsis*. The T-DNA mutants did not show any significant phenotype differences as compared with the *Col-0* wild-type plant population. The *Arabidopsis* genes need to be characterised in more detail. Work is also necessary to determine what functions they have normally in the plant and whether these are similar or different from their roles in other organisms.

4.4.5 Possible limitations of my methods

The cell size measurements performed on the T-DNA insertion mutants were done with a PC by analysing pictures taken with a camera. It is therefore possible that this mode of operation introduced some biases and could be seen as a limitation for a proper interpretation of the results. Both plant size and perimeter or the area measurements were performed on macroscopic pictures taken above the plants. Although pictures were systematically taken using the same camera shooting angle I, however, introduced some modifications to correct the angles of different viewpoints as stated in Chapter 4.2.4.2 .

In the analysis of plants growing on sterile medium I was using young plants with a limited vertical dimension and it is unlikely that photos from the top have introduced biases.

Using a PC, I restricted the analysis to the methods and protocols described in Section 4.2.6 . However, some alterations could have been introduced if the leaf number was not selected carefully or the position of the cut area was not identified precisely.

Furthermore, all the analyses were performed by me and the technical variability during handling of the material and the analysis of the results should be considered minimal.

4.4.6 Possible alternative methods

First of all, the ultimate experiment would be to perform a gene editing using CRISPR/Cas9 to trigger a mutagenesis in the gene of interest. This could have given me the chance to test the *Pfd4* mutants of the Prefoldin complex, where I was not able to produce a homozygous mutant.

Other alternate methods could be to use alternative measurements for cell size. Either other applications than ImageJ or to use more sophisticated microscopical techniques in 3D for instance using a collaboration with research groups experts in this field of research.

Chapter 5

Conclusions

and

Future Outlook

5.1 Conclusions

In this Thesis, I showed how to build a tool to create a comprehensive outlook of proteins that are regulating cell size. Cell size can be regulated in many ways as I have showed in the introduction part. There are three models that can be used: sizer, adder and timer (Tzur et al., 2009). In this thesis I was looking for cell size regulators that have an effect to the overall cell size without investigating the actual mechanisms behind the regulation.

5.1.1 Summarizing the results

In the Results chapters of my PhD Thesis, I described how I built two bioinformatics tools to predict proteins that control cell size. Then I performed experiments on *Arabidopsis* and Human T-Cells to determine whether some of the proteins predicted by my bioinformatics analyses affect cell size of those species.

5.1.2 Bioinformatics analyses

My first aim was to create a bioinformatics tool that can identify novel cell size orthologs in *Homo sapiens* and *Arabidopsis*. I successfully achieved this aim because I created a tool that collects the known cell size information then gives predictions for experimental testing. I made this tool publicly available (`http://www.orthologfindertool.com`). With the help of this tool, I exported a pool of candidates described in detail in Chaper2, then I selected 17 of them to test experimentally after a 9-step narrowing process. The 9-step narrowing process is listed in Section 2.3.2

5.1.3 Predicted candidates of the bioinformatic tool

As I showed in Section from these 17 candidates, I selected 5 proteins plus the whole prefoldin complex. I tested these candidates experimentally, for the results see Table 5.1-1.

Table 5.1-1: Summary of the experimental tests

In this table I show the results of the experimental tests in *Arabidopsis* and human T-Cells. I tested all of the candidates listed in the table for *Arabidopsis* and all but two in human T-Cells. I did not list VPS18 and PFDN4 that I did not tested for various reasons in these organisms.

A. thaliana	*Test / Result*	*H. sapiens*	*Test / Result*
At4g16970	Yes / Negative	CDC7	Not tested
At2g07340	Yes / Negative	PFD1	Not tested
At3g22480	Yes / Negative	PFD2	Not tested
At5g49510	Yes / Positive	PFD3/VBP1	Yes / Negative
At5g23290	Yes / Negative	PFD5	Not successful
At1g29990	Yes / Positive	PFD6	Not successful
Tctp	Yes / Negative	TCTP	Yes / Positive
Mos14	Yes / Negative	TNPO3	Not tested
At5g26110	Yes / Negative	TP53RK	Not successful

5.1.4 Investigation of cell size regulators in human T-cells and Arabidopsis

As shown in Table 5.1-1, I tested experimentally the 6 predicted candidate proteins based on my bioinformatics tool using human T-Cells. The experiments are based on reducing the induction of a given protein during the G_0 to G_1 transition using siRNA. Three of these experiments were successful for TCTP, VBP1 and PFDN6 proteins, while the PFDN2, PFDN5 and TP53RK protein results were unsuccessful, as I was not able to verify the reduction of the protein levels by Western Blotting.

In *Arabidopsis*, I tested whether genetic knockouts of 10 candidates have effects on cell or plant size. I tested all of them as listed in the Table 5.1-1. Only two of them were known to affect cell size earlier: the *Pfd6* and *Tctp* gene. The others have been never reported to affect or regulate cell or plant size. I tested the role in Arabidopsis cell size and organ growth for all of the 10 gene candidates using knock-out mutant lines.

5.1.5 TCTP/TPT1

Reducing the induction of one of TCTP caused a significant difference in cell size in human T-Cells. TCTP is a versatile protein that is involved in many fundamental biological processes and disorders either in human or other species (Bommer et al., 2017). Possible it can be involved in cell size control in human, but interestingly it has been never reported earlier. If my results stand for TCTP in further repeated experiments it could make a higher impact as the protein is known to be involved in many cancer types (Telerman & Amson, 2009). Novel publications citing that TCTP is induces proliferation and cell growth in early stages of colorectal cancer (Xiao et al., 2016). As well as some already cite TCTP as a therapeutic target in some cancer types such as melanoma (Boia-Ferreira et al., 2017). Despite my results showed significant difference there is a need for more experiments to test my findings and conclude whether TCTP protein is involved in cell size control.

In *Arabidopsis*, the homozygous T-DNA insertion mutant *tctp1* (Knock-out of gene At3g16640) did not result in any significant difference in phenotype compared to the wild-type - Col-0 plants. Previously, *tctp1* showed a severe delay in development and critical growth defects with small organs and short plant stature (Brioudes et al., 2010). Moreover, Berkowitz *et al.* (2008) also described *tctp1* mutants with an impaired vegetative devolvement and a leaf expansion that was slowed down due to reduced cell size. I was unable to see similar phenotypes despite I run at least 3 complete independent experiments on different SALK mutants.

5.1.5.1 MCM7

As I showed, I used MCM7 protein knock-downs as a technical control for cell cycle. I also checked the cell size of Mcm7 transfected cells. Importantly, transfecting siRNA to reduce the expression of Mcm7, caused a reduction of cell size in G_1 and S-phases. It should be noted that gene knockouts of *MCM* genes are lethal and the effect of reducing Mcm7 expression on cell size has not been investigated previously. As this finding suggests that the MCM complex as a whole can reduce cell size, I tested MCM4 knock-downs that showed a very similar reduction pattern of cell size distribution. MCM protein knock-downs known to be arrested in the G_0 phase of the cell cycle (Orr et al., 2010). It would be interesting to check whether the effect of other members of the MCM complex show a similar pattern.

I did not investigate the effect of Mcm7 (At4g02060) knock-out in *Arabidopsis*, since a mutation in this gene is embryo lethal as it is known to be essential for early *Arabidopsis* development (Springer et al., 2000). Moreover, it is also known if we reduce MCM levels using endosperm-specific RNAi constructs it results in the up-regulation of DNA repair transcripts (Herridge et al., 2014).

5.1.6 PFD6/PFDN6

Pfd6 has been described to affect cell size and reduce plant size in *Arabidopsis*, while affecting the arrangement of microtubules (Gu et al., 2008). I was using *pfd6* as a positive control of my experiments and my results corroborate the findings of Gu et al in 2008.

In human T-Cells, I tested the knock-down effect of PFDN6, however there were no significant difference between the knock-down and control populations. The reason behind could be it does not fulfil any essential chaperonin function in human or other proteins surpasses the effect of the PFDN6 protein.

5.1.7 VBP1/PFD3

Based on the experiments performed in *Arabidopsis*, I gave suggestions toward that one of the prefoldin orthologs, namely *Pfd3* is a cell size regulator. I have showed that *Pfd3* mutants have significantly smaller cell size. Moreover, *Pfd3* knockout plants showed significantly larger seeds and longer siliques compared with the control Col-0 samples. This kind of phenotype has never been described before for *pfd3* mutants. As my results show that *pfd3* T-DNA mutant has significantly larger seeds. This larger seed phenotype has been reported previously for *da1* gene overexpression (Li et al., 2008). Furthermore, a very similar change in seed phenotype has been recently reported using the loss-of-function mutant of *raptor1b*, which functions in the TOR pathway (Salem et al., 2017). Producing larger seeds is potentially important for food products and further work is needed to understand the mechanisms that control seed size, particularly of agriculturally important plants.

In human T-Cells, I tested the knock-down effect of VBP1, however there were no significant difference between the knock-down and control population of cells.

5.1.8 TNPO3/Mos14

I also tested the *Mos14* gene knock-out (KO) plants and checked whether they affect cell size or plant development in *Arabidopsis*. For *Mos14* I found that the size of the cotyledons and leaf #1 are significantly smaller (< 0.05) than the Col-0 plants. This has never been reported. The reason behind the size difference in leaves could be the change in nuclear transport (Xu et al., 2011). It is know that *Mos14* is needed for plant immunity through proper splicing of R genes (Tamura & Hara-Nishimura, 2014). Reduction of the size of roots and lateral organs in *Arabidopsis* has been reported to other transport molecules like HASTY (Bollman, 2003), that would explain the changes of size of *Mos14* KO mutants. Despite I selected, I did not test the TNPO3 knock-downs in human T-Cells due to lack of time.

5.1.9 Negative Results

As I have stated above that, I experienced verified negative results in human T-Cells for VBP1 and PFDN6 protein siRNA knock-downs. While, in *Arabidopsis* I did not observe any significant difference in size or growth for *Cdc7* (At4g16970) and *tp53rk* (At5g26110), as well as the other three members of the prefoldin complex that I tested *pfd1* (At2g07340) *pfd2* (At3g22480), and *pfd5* (At5g23290) compare to the wild-type Col-0.

5.1.10 Unsuccessful experiments

On Table 5.1-1, it can be seen that some of my experiments did not work in human T-Cells. The experiments for PFDN2 and PFDN5 siRNA transfected human T-Cells were unsuccessful. During my PhD studies, I tried to overcome this issue in several ways by changing the siRNA manufacturer. I tried either the siRNA provided by Dharmacon (Table 3.2-9) or the custom made siRNA from SigmaAldrich (Table 3.2-10), none of them produced a knock-down sufficient enough to reduce the protein level as tested by Western blot analyses. Moreover, I tried unsuccessfully to double the amount of the transfecting siRNA that did not shows well any reduction in the protein levels. I also had similar issues with TP53RK experiments in human T-Cells.

5.1.11 The GOOT Tool

As a follow-up of the orthology analysis of cell size regulators, I have created a generalised bioinformatic tool. This tool goes further and extends the tool that I created earlier to predict the conserved orthologous groups for any biological functions that are listed in the GO Slim database. I made this tool also publicly available at http://go.orthologfindertool.com. I have shown a couple of examples which illustrate that it can indeed identify which conserved orthologous groups are responsible for a common GO function. For example, I showed that this tool is good to identify novel annotations for the GO term "*GO:0051726 - regulation of cell cycle".* As well as I tested this tool for more GO term that I did not included in this thesis such as "*GO:0006629 - lipid metabolic process*" and "*GO:0044281 - small molecule metabolic process*".

Furthermore, I showed that this tool can help to extend the GO annotations inside the Gene Ontology Database, as it can highlight if a conserved function is missing from the annotations of proteins that are associated with an otherwise highly conserved functional orthologous group. Currently I'm testing this tool with more Gene Ontology term beyond GO-SLIM set.

5.2 Future Outlook

5.2.1 Bioinformatics

5.2.1.1 Validation of my tools

As I have shown in Chapter 2, I was able to predict candidate proteins with specific functions, most of which now need to be investigated further in large-scale wet-laboratory experiments. Further, predictions using my tool have not been compared with those of other tools. A systematic comparison of the predictions obtained with my tool as compared with others is therefore required.

5.2.1.1.1 Validating the Ortholog Finder Tool

My tool is for predicting functional orthologs in human and *Arabidopsis*. As a first option I would employ a method to check the BLAST suite (Götz et al., 2008). I would enter a list of proteins from my initial size-control list in line with a set of current predictions in human to the BLAST2Go tool (Götz et al., 2008), then determine how many of them are predicted in common by both tools. The BLAST2Go suite is a good combination of various strategies since it contains general sequence management features and it has high-throughput capabilities (Götz et al., 2008). I favor a scenario in which the predictions made by my tool and those made by BLAST2GO would overlap but more candidates would be made available using my tool. I would also use a second method called RAST (Rapid Annotation of microbial genomes using Subsystems Technology) to test the quality of my predictions (Overbeek et al., 2014). The RAST method uses SEED, which contains consistent and accurate genome annotations and the annotations are scalable in more than 60,000 genomes. (Overbeek et al., 2014).

5.2.1.1.2 Validating GOOT

There are approaches to extend Gene Ontology based on homology as well as with phylogenetic annotation (Gaudet et al., 2011). I would test my predictions using the PAINT tool. I would pick 100 orthology groups, where there were 7 proteins in seven species, although all but one species has been annotated to a specific function. I would give 6 proteins - that have already been annotated as the sources and test the tool. If it gives back a prediction for the 7th organism, the protein that I have predicted in most of the cases (over 80%) it would be a good source for predicting protein candidates.

5.2.1.2 Predictions for less characterized species

The generalised **OrthologFinder Tool** can identify conserved proteins responsible for any GO-slim annotated biological functions. It will be interesting to see if the tool can be expanded to predict GO annotations for poorly characterised species, just by the presence of genes from these species in eggNOG groups. Sequencing of uncharacterized genomes is still a hot field in scientific research and as of 2013 more than 644 fully sequenced eukaryote genomes are in NCBI databases (Ellegren, 2014), while as of 2015 more than 20,000 complete prokaryote genomes have been sequenced (Land et al., 2015). Newly sequenced genomes contain uncharacterized proteins and while proteomics in non-model organisms has flourished in the recent years, researchers still try to add characterisation based on sequence homology (Armengaud et al., 2014). If we take a look at the goals and maintenance of the eggNOG database, they are regularly including data from non-model organisms and expanding the database with similar goals (Huerta-Cepas et al., 2016). As of version 4.5.1, the database contains 2,031 organisms with more than 190,000 orthologous groups (Figure 5.2-1). Based on eggNOG data my tool would be able to predict novel functional GeneOntology annotations for non-model organisms based on model organisms' data. This could be used together with existing tools that try to predict GO annotations based on sequence similarity or other features (Mitchell et al., 2015).

Figure 5.2-1: eggNOG database version 4.5.1

The database contains 2,031 organisms, 352 viruses and more than 190,000 orthologous groups.

5.2.1.3 Experimental validation

If time and money would permit me, I would test experimentally all of my 17 candidate proteins (described in Section 2.3.2). It would be good to see how many of them would have the predicted function either in *Arabidopsis* or human. I would like to run experiments with cells isolated from more human tissues to determine whether the protein has similar effects on the size of a wide variety of cells or just specific cell types. Further experiments could be done on mice in which conditional gene knock-out experiments (whole mouse or cell type-specific) could be performed. To test the effects of the predictions in less-curated species I would base my sources on known organisms and set of proteins, predict in less-curated but fully sequenced organisms, then test them experimentally.

5.2.2 Arabidopsis

Carrying out experiments on *Arabidopsis* is labour-intensive, and I was only able to test 10 candidates during my PhD. If more technical resources were available, it would be good to test a much larger set of targets to determine which affect cell size as well as the size of seeds and other plant structures. Such experiments are important scientifically, but the results are also of potential interest for agriculturally important crops.

An elegant way to show a complementation effect due to gene redundancy in the prefoldin gene family in *Arabidopsis* is to create double or triple mutants using all combinations of *pfd1*x*pfd2, pfd1*x*pfd3 or pfd1*x*pfd2xpfd3.* There are 2,400 gene knock-out mutants with significant phenotypes in *Arabidopsis* (Lloyd & Meinke, 2012) and there are more than 27,000 genes in the genome (Swarbreck et al., 2008). While there could be several reasons to explain why only 10% of the gene knockouts cause a change in phenotypes my hypothesis is that *Arabidopsis* encodes some gene families with members having redundant functions. There are no seed repositories for double or triple mutants except for donations from special projects such as the MAPK initiative that deposited 275 double mutants to ABRC (Su & Krysan, 2016). Generating these double and triple mutants of the genes encoding prefoldins would require manual crosses but could potentially provide interesting results.

5.2.3 Human

As is the case for extending the coverage of *Arabidopsis* experiments, it would be possible to carry out high-throughput siRNA screens to identify other proteins that regulate the size of human T-Cells. Other experimental approaches are also possible. In recent years enormous efforts have been made to modify and manipulate the DNA of a given T-Cell genetically. These efforts were unsuccessful for a long time, but finally, Schuman et al. (2015) were able to successfully use CRISPR/Cas9 technology in human T-Cells (Schumann et al., 2015). They deleted the gene encoding the CXCR4 cell-surface receptor protein, which is a co-receptor for HIV, using electroporation of CRISPR/Cas9 ribonucleoproteins (Schumann et al., 2015). Unfortunately, this method was not available when I started my PhD in 2013. However, it opens the door for future generations in T-Cell related research into the requirement for genes predicted by my bioinformatics analyses to be involved in cell size control. I note that this approach is different from reducing expression with siRNA. Mcm7 would not be identified by a CRISPR/Cas9 knockout, whereas it is using a siRNA method

Chapter 6

References

and

Supplementary Information

6.1 References

Abe, A., Takahashi-Niki, K., Takekoshi, Y., Shimizu, T., Kitaura, H., Maita, H., ... Ariga, H. (2013). Prefoldin plays a role as a clearance factor in preventing proteasome inhibitor-induced protein aggregation. *Journal of Biological Chemistry*, *288*(39), 27764–27776. https://doi.org/10.1074/jbc.M113.476358

Abe, Y., Matsumoto, S., Wei, S., Nezu, K., Miyoshi, A., Kito, K., ... Enomoto, Y. (2001). Cloning and Characterization of a p53-related Protein Kinase Expressed in Interleukin-2-activated Cytotoxic T-cells, Epithelial Tumor Cell Lines, and the Testes. *Journal of Biological Chemistry*, *276*(47), 44003–44011. https://doi.org/10.1074/jbc.M105669200

Achard, P., Liao, L., Jiang, C., Desnos, T., Bartlett, J., Fu, X., & Harberd, N. P. (2007). DELLAs Contribute to Plant Photomorphogenesis. *PLANT PHYSIOLOGY*, *143*(3), 1163–1172. https://doi.org/10.1104/pp.106.092254

Acunzo, J., Baylot, V., So, A., & Rocchi, P. (2014). TCTP as therapeutic target in cancers. *Cancer Treatment Reviews*, *40*(6), 760–769. https://doi.org/10.1016/j.ctrv.2014.02.007

Acuto, O., & Michel, F. (2003). CD28-mediated co-stimulation: a quantitative support for TCR signalling. *Nature Reviews Immunology*, *3*(12), 939–951. https://doi.org/10.1038/nri1248

Aichinger, E., Kornet, N., Friedrich, T., & Laux, T. (2012). Plant stem cell niches. *Annual Review of Plant Biology*, *63*, 615–36. https://doi.org/10.1146/annurev-arplant-042811-105555

Ajchenbaum, F., Ando, K., DeCaprio, J. A., & Griffin, J. D. (1993). Independent regulation of human D-type cyclin gene expression during G1 phase in primary human T lymphocytes. *The Journal of Biological Chemistry*, *268*(6), 4113–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8382693

Aldea, M., Jenkins, K., & Csikász-Nagy, A. (2017). Growth Rate as a Direct Regulator of the Start Network to Set Cell Size. *Frontiers in Cell and Developmental Biology*, *5*(May), 1–6. https://doi.org/10.3389/fcell.2017.00057

Allemand, E., Dokudovskaya, S., Bordonné, R., & Tazi, J. (2002). A conserved Drosophila transportin-serine/arginine-rich (SR) protein permits nuclear import of Drosophila SR protein splicing factors and their antagonist repressor splicing factor 1. *Molecular Biology of the Cell*, *13*(7), 2436–47. https://doi.org/10.1091/mbc.E02-02-0102

Allison, K. E., Coomber, B. L., & Bridle, B. W. (2017). Metabolic reprogramming in the tumour microenvironment: a hallmark shared by cancer cells and T lymphocytes. *Immunology*, *152*(2), 175–184. https://doi.org/10.1111/imm.12777

Alonso-Blanco, C., Blankestijn-de Vries, H., Hanhart, C. J., & Koornneef, M. (1999). Natural allelic variation at seed size loci in relation to other life history traits of Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(8), 4710–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10200327

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of

protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–402. https://doi.org/10.1093/nar/25.17.3389

Amson, R., Pece, S., Lespagnol, A., Vyas, R., Mazzarol, G., Tosoni, D., ... Telerman, A. (2012). Reciprocal repression between P53 and TCTP. *Nature Medicine*, *18*(1), 91–99. https://doi.org/10.1038/nm.2546

Andree, H., Thiele, H., Fähling, M., Schmidt, I., & Thiele, B.-J. (2006). Expression of the human TPT1 gene coding for translationally controlled tumor protein (TCTP) is regulated by CREB transcription factors. *Gene*, *380*(2), 95–103. https://doi.org/10.1016/j.gene.2006.05.018

Anfinsen, C. B. (1973). Principles that Govern the Folding of Protein Chains. *Science*, *181*(4096), 223–230. https://doi.org/10.1126/science.181.4096.223

Appleman, L. J., van Puijenbroek, A. A. F. L., Shu, K. M., Nadler, L. M., & Boussiotis, V. A. (2002). CD28 costimulation mediates down-regulation of p27kip1 and cell cycle progression by activation of the PI3K/PKB signaling pathway in primary human T cells. *Journal of Immunology (Baltimore, Md. : 1950)*, *168*(6), 2729–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11884439

Armengaud, J., Trapp, J., Pible, O., Geffard, O., Chaumot, A., & Hartmann, E. M. (2014). Non-model organisms, a species endangered by proteogenomics. *Journal of Proteomics*, *105*, 5–18. https://doi.org/10.1016/j.jprot.2014.01.007

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, *25*(1), 25–9. https://doi.org/10.1038/75556

Ashcroft, M., Kubbutat, M. H., & Vousden, K. H. (1999). Regulation of p53 function and stability by phosphorylation. *Molecular and Cellular Biology*, *19*(3), 1751–8. https://doi.org/10.1128/MCB.19.3.1751

Averous, J., & Proud, C. G. (2006). When translation meets transformation: the mTOR story. *Oncogene*, *25*(48), 6423–6435. https://doi.org/10.1038/sj.onc.1209887

Bae, S.-Y., Kim, H. J., Lee, K.-J., & Lee, K. (2015). Translationally Controlled Tumor Protein induces epithelial to mesenchymal transition and promotes cell migration, invasion and metastasis. *Scientific Reports*, *5*(1), 8061. https://doi.org/10.1038/srep08061

Baserga, R. (2007). Is Cell Size Important ? *Cell Cycle*, *6*(7), 814–816.

Bazile, F., Pascal, A., Arnal, I., Le Clainche, C., Chesne, F., & Kubiak, J. Z. (2009). Complex relationship between TCTP, microtubules and actin microfilaments regulates cell shape in normal and cancer cells. *Carcinogenesis*, *30*(4), 555–565. https://doi.org/10.1093/carcin/bgp022

Beltrao, P., Ryan, C., & Krogan, N. J. (2012). Comparative interaction networks: bridging genotype to phenotype. *Advances in Experimental Medicine and Biology*, *751*(December 1831), 139–56. https://doi.org/10.1007/978-1-4614-3567-9_7

Benndorf, R., Nürnberg, P., & Bielka, H. (1988). Growth phase-dependent proteins of the Ehrlich ascites tumor analyzed by one- and two-dimensional electrophoresis. *Experimental Cell Research*, *174*(1), 130–138. https://doi.org/10.1016/0014-4827(88)90148-6

Berciano, M. T., Novell, M., Villagra, N. T., Casafont, I., Bengoechea, R., Val-Bernal, J. F., & Lafarga, M. (2007). Cajal body number and nucleolar size correlate with the cell body mass in human sensory ganglia neurons. *Journal of Structural Biology*,

158(3), 410–420. https://doi.org/10.1016/j.jsb.2006.12.008

Berkowitz, O., Jost, R., Pollmann, S., & Masle, J. (2008). Characterization of TCTP, the Translationally Controlled Tumor Protein, from Arabidopsis thaliana. *The Plant Cell Online*, *20*(12), 3430–3447. https://doi.org/10.1105/tpc.108.061010

Besson, A., Dowdy, S. F., & Roberts, J. M. (2008). CDK Inhibitors: Cell Cycle Regulators and Beyond. *Developmental Cell*, *14*(2), 159–169. https://doi.org/10.1016/j.devcel.2008.01.013

Betsch, L., Savarin, J., Bendahmane, M., & Szecsi, J. (2017). Roles of the Translationally Controlled Tumor Protein (TCTP) in Plant Development. *Results and Problems in Cell Differentiation*, *64*, 149–172. https://doi.org/10.1007/978-3-319-67591-6_7

Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., ... Canaider, S. (2013). An estimation of the number of cells in the human body. *Annals of Human Biology*, *40*(6), 463–471. https://doi.org/10.3109/03014460.2013.807878

Bieluszewski, T., Galganski, L., Sura, W., Bieluszewska, A., Abram, M., Ludwikow, A., ... Sadowski, J. (2015). AtEAF1 is a potential platform protein for Arabidopsis NuA4 acetyltransferase complex. *BMC Plant Biology*, *15*(1), 1–15. https://doi.org/10.1186/s12870-015-0461-1

Bierer, B. E., & Burakoff, S. J. (1988). T cell adhesion molecules. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *2*(10), 2584–90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2838364

Björklund, M., Taipale, M., Varjosalo, M., Saharinen, J., Lahdenperä, J., & Taipale, J. (2006). Identification of pathways regulating cell size and cell-cycle progression by RNAi. *Nature*, *439*(7079), 1009–13. https://doi.org/10.1038/nature04469

Blow, J. J., & Hodgson, B. (2002). Replication licensing--defining the proliferative state? *Trends in Cell Biology*, *12*(2), 72–8. https://doi.org/10.1016/S0962-8924(01)02203-6

Böhm, H., Benndorf, R., Gaestel, M., Gross, B., Nürnberg, P., Kraft, R., ... Bielka, H. (1989). The growth-related protein P23 of the Ehrlich ascites tumor: translational control, cloning and primary structure. *Biochemistry International*, *19*(2), 277–86. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2479380

Böhni, R., Riesgo-Escovar, J., Oldham, S., Brogiolo, W., Stocker, H., Andruss, B. F., ... Hafen, E. (1999). Autonomous control of cell and organ size by CHICO, a Drosophila homolog of vertebrate IRS1-4. *Cell*, *97*(7), 865–875. https://doi.org/10.1016/S0092-8674(00)80799-0

Boia-Ferreira, M., Basílio, A. B., Hamasaki, A. E., Matsubara, F. H., Appel, M. H., Da Costa, C. R. V., ... Senff-Ribeiro, A. (2017). TCTP as a therapeutic target in melanoma treatment. *British Journal of Cancer*, *117*(5), 656–665. https://doi.org/10.1038/bjc.2017.230

Bollman, K. M. (2003). HASTY, the Arabidopsis ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. *Development*, *130*(8), 1493–1504. https://doi.org/10.1242/dev.00362

Bommer, U.-A. (2017). The Translational Controlled Tumour Protein TCTP: Biological Functions and Regulation. In *Results and problems in cell differentiation* (Vol. 64, pp. 69–126). https://doi.org/10.1007/978-3-319-67591-6_4

Bommer, U.-A., & Thiele, B.-J. (2004). The translationally controlled tumour protein

(TCTP). *The International Journal of Biochemistry & Cell Biology*, *36*(3), 379–85. https://doi.org/10.1016/S1357-2725(03)00213-9

Bommer, U. A., Iadevaia, V., Chen, J., Knoch, B., Engel, M., & Proud, C. G. (2015). Growth-factor dependent expression of the translationally controlled tumour protein TCTP is regulated through the PI3-K/Akt/mTORC1 signalling pathway. *Cellular Signalling*, *27*(8), 1557–1568. https://doi.org/10.1016/j.cellsig.2015.04.011

Bommer, U. A., Vine, K. L., Puri, P., Engel, M., Belfiore, L., Fildes, K., ... Aghmesheh, M. (2017). Translationally controlled tumour protein TCTP is induced early in human colorectal tumours and contributes to the resistance of HCT116 colon cancer cells to 5-FU and oxaliplatin. *Cell Communication and Signaling*, *15*(1), 1–15. https://doi.org/10.1186/s12964-017-0164-3

Bonnevier, J. L., & Mueller, D. L. (2002). Cutting edge: B7/CD28 interactions regulate cell cycle progression independent of the strength of TCR signaling. *Journal of Immunology (Baltimore, Md.: 1950)*, *169*(12), 6659–63. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12471093

Boomer, J. S., & Green, J. M. (2010). An enigmatic tail of CD28 signaling. *Cold Spring Harbor Perspectives in Biology*, *2*(8), a002436. https://doi.org/10.1101/cshperspect.a002436

Boonen, G. J., van Dijk, A. M., Verdonck, L. F., van Lier, R. A., Rijksen, G., & Medema, R. H. (1999). CD28 induces cell cycle progression by IL-2-independent down-regulation of p27kip1 expression in human peripheral T lymphocytes. *Eur J Immunol*, *29*(3), 789–798. https://doi.org/10.1002/(SICI)1521-4141(199903)29:03<789::AID-IMMU789>3.0.CO;2-5

Boyd, R. L., Tucek, C. L., Godfrey, D. I., Izon, D. J., Wilson, T. J., Davidson, N. J., ... Hugo, P. (1993). The thymic microenvironment. *Immunology Today*, *14*(9), 445–459. https://doi.org/10.1016/0167-5699(93)90248-J

Bradford, Y., Conlin, T., Dunn, N., Fashena, D., Frazer, K., Howe, D. G., ... Westerfield, M. (2011). ZFIN: Enhancements and updates to the zebrafish model organism database. *Nucleic Acids Research*, *39*(SUPPL. 1), 822–829. https://doi.org/10.1093/nar/gkq1077

Braun, D. A., Rao, J., Mollet, G., Schapiro, D., Daugeron, M. C., Tan, W., ... Hildebrandt, F. (2017). Mutations in KEOPS-complex genes cause nephritic syndrome with primary microcephaly. *Nature Genetics*, *49*(10), 1529–1538. https://doi.org/10.1038/ng.3933

Brioudes, F., Thierry, A.-M., Chambrier, P., Mollereau, B., & Bendahmane, M. (2010). Translationally controlled tumor protein is a conserved mitotic growth integrator in animals and plants. *Proceedings of the National Academy of Sciences*, *107*(37), 16384–16389. https://doi.org/10.1073/pnas.1007926107

Brown, E. J., Albers, M. W., Shin, T. B., Ichikawa, K., Keith, C. T., Lane, W. S., & Schreiber, S. L. (1994). A mammalian protein targeted by G1-arresting rapamycin-receptor complex. *Nature*, *369*(6483), 756–8. https://doi.org/10.1038/369756a0

Brown, M., & Wittwer, C. (2000). Flow cytometry: Principles and clinical applications in hematology. *Clinical Chemistry*, *46*(8 II), 1221–1229. https://doi.org/1530-8561

Brownlie, R. J., & Zamoyska, R. (2013). T cell receptor signalling networks: branched,

diversified and bounded. *Nature Reviews Immunology*, *13*(4), 257–269. https://doi.org/10.1038/nri3403

Bruckner, F. P., Xavier, A. D. S., Cascardo, R. D. S., Otoni, W. C., Zerbini, F. M., & Alfenas-Zerbini, P. (2017). Translationally controlled tumour protein (TCTP) from tomato and Nicotiana benthamiana is necessary for successful infection by a potyvirus. *Molecular Plant Pathology*, *18*(5), 672–683. https://doi.org/10.1111/mpp.12426

Buck, M. D., O'Sullivan, D., & Pearce, E. L. (2015). T cell metabolism drives immunity. *The Journal of Experimental Medicine*, *212*(9), 1345–1360. https://doi.org/10.1084/jem.20151159

Buck, M. D., Sowell, R. T., Kaech, S. M., & Pearce, E. L. (2017). Metabolic Instruction of Immunity. *Cell*, *169*(4), 570–586. https://doi.org/10.1016/j.cell.2017.04.004

Call, M. E., & Wucherpfennig, K. W. (2007). Common themes in the assembly and architecture of activating immune receptors. *Nature Reviews. Immunology*, *7*(11), 841–50. https://doi.org/10.1038/nri2186

Camon, E. (2004). The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Research*, *32*(90001), 262D–266. https://doi.org/10.1093/nar/gkh021

Campisi, J. (1996). Replicative senescence: an old wives tale. *Cell*, *84*, 497–500. https://doi.org/10.1016/S0092-8674(00)81023-5

Cantrell, D. A., Collins, M. K., & Crumpton, M. J. (1988). Autocrine regulation of T-lymphocyte proliferation: differential induction of IL-2 and IL-2 receptor. *Immunology*, *65*(3), 343–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3264805

Cantrell, D. A., & Smith, K. A. (1983). Transient expression of interleukin 2 receptors. Consequences for T cell growth. *The Journal of Experimental Medicine*, *158*(6), 1895–911. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6606011

Cantrell, D. A., & Smith, K. A. (1984). The Interleukin-2 T-Cell System : A New Cell Growth Model Author. *Science*, *224*(4655), 1312–1316.

Cao, B., Lu, Y., Chen, G., & Lei, J. (2010). Functional characterization of the translationally controlled tumor protein (TCTP) gene associated with growth and defense response in cabbage. *Plant Cell, Tissue and Organ Culture (PCTOC)*, *103*(2), 217–226. https://doi.org/10.1007/s11240-010-9769-6

Cao, J. (2016). Analysis of the Prefoldin Gene Family in 14 Plant Species. *Frontiers in Plant Science*, *7*(March), 1–14. https://doi.org/10.3389/fpls.2016.00317

Cao, S., Carlesso, G., Osipovich, A. B., Llanes, J., Lin, Q., Hoek, K. L., ... Ruley, H. E. (2008). Subunit 1 of the prefoldin chaperone complex is required for lymphocyte development and function. *Journal of Immunology*, *181*(1), 476–84. https://doi.org/10.4049/jimmunol.181.1.476

Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., ... Gaudet, P. (2009). AmiGO: Online access to ontology and annotation data. *Bioinformatics*, *25*(2), 288–289. https://doi.org/10.1093/bioinformatics/btn615

Castellano, M. M., & Sablowski, R. (2008). Phosducin-Like Protein 3 is required for microtubule-dependent steps of cell division but not for meristem growth in Arabidopsis. *The Plant Cell*, *20*(4), 969–981. https://doi.org/10.1105/tpc.107.057737

Cecchini, S., Virag, T., & Kotin, R. M. (2011). Reproducible High Yields of Recombinant Adeno-Associated Virus Produced Using Invertebrate Cells in 0.02- to 200-Liter Cultures. *Human Gene Therapy*, *22*(8), 1021–1030. https://doi.org/10.1089/hum.2010.250

Cereda, M., Mourikis, T. P., & Ciccarelli, F. D. (2016). Genetic Redundancy, Functional Compensation, and Cancer Vulnerability. *Trends in Cancer*, *2*(4), 160–162. https://doi.org/10.1016/j.trecan.2016.03.003

Chan, T. H. M., Chen, L., Liu, M., Hu, L., Zheng, B. J., Poon, V. K. M., ... Guan, X. Y. (2012). Translationally controlled tumor protein induces mitotic defects and chromosome missegregation in hepatocellular carcinoma development. *Hepatology*, *55*(2), 491–505. https://doi.org/10.1002/hep.24709

Chan, Y.-H. M., & Marshall, W. F. (2012). How Cells Know the Size of Their Organelles. *Science*, *337*(6099), 1186–1189. https://doi.org/10.1126/science.1223539

Charollais, R. H., Tiwari, S., & Thomas, N. S. (1994). Into and out of G1: the control of cell proliferation. *Biochimie*, *76*(9), 887–94. https://doi.org/10.1016/0300-9084(94)90191-0

Chen, D., Wilkinson, C. R. M., Watt, S., Penkett, C. J., Toone, W. M., Jones, N., & Bähler, J. (2007). High-Resolution Crystal Structure and In Vivo Function of a Kinesin-2 Homologue in Giardia intestinalis. *Molecular Biology of the Cell*, *19*(1), 308–317. https://doi.org/10.1091/mbc.E07

Chen, Y., Fujita, T., Zhang, D., Doan, H., Pinkaew, D., Liu, Z., ... Fujise, K. (2011). Physical and functional antagonism between tumor suppressor protein p53 and fortilin, an anti-apoptotic protein. *Journal of Biological Chemistry*, *286*(37), 32575–32585. https://doi.org/10.1074/jbc.M110.217836

Cheng, X., Li, J., Deng, J., Li, Z., Meng, S., & Wang, H. (2012). Translationally controlled tumor protein (TCTP) downregulates Oct4 expression in mouse pluripotent cells. *BMB Reports*, *45*(1), 20–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22281008

Chica, N., Rozalén, A. E., Pérez-Hidalgo, L., Rubio, A., Novak, B., & Moreno, S. (2016). Nutritional control of cell size by the greatwall-endosulfine-PP2A·B55 pathway. *Current Biology*, *26*(3), 319–330. https://doi.org/10.1016/j.cub.2015.12.035

Chitpatima, S. T., Makrides, S., Bandyopadhyay, R., & Brawerman, G. (1988). Nucleotide sequence of a major messenger RNA for a 21 kilodalton polypeptide that is under translational control in mouse tumor cells. *Nucleic Acids Research*, *16*(5), 2350. https://doi.org/10.1093/nar/16.5.2350

Chu, Z.-H., Liu, L., Zheng, C.-X., Lai, W., Li, S.-F., Wu, H., ... Guan, Y.-F. (2011). Proteomic analysis identifies translationally controlled tumor protein as a mediator of phosphatase of regenerating liver-3-promoted proliferation, migration and invasion in human colon cancer cells. *Chinese Medical Journal*, *124*(22), 3778–85. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22340241

Clough, S. J., & Bent, A. F. (1998). Floral dip: A simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant Journal*, *16*(6), 735–743. https://doi.org/10.1046/j.1365-313X.1998.00343.x

Cobrinik, D. (2005). Pocket proteins and cell cycle control. *Oncogene*, *24*(17), 2796–2809. https://doi.org/10.1038/sj.onc.1208619

Conlon, I. J., Dunn, G. A., Mudge, A. W., Raff, M. C., Muscle, M. R. C., Unit, C. M., ... Se, L. (2001). Extracellular control of cell size. *Nature Cell Biology*, *3*(10), 918–921. https://doi.org/10.1038/ncb1001-918

Conlon, I., & Raff, M. (1999). Size Control in Animal Development Review. *Cell*, *96*, 235–244. https://doi.org/10.1016/S0092-8674(00)80563-2

Conlon, I., & Raff, M. (2003). Differences in the way a mammalian cell and yeast cells coordinate cell growth and cell-cycle progression. *Journal of Biology*, *2*(7), 1–10. https://doi.org/10.1186/1475-4924-2-7

Cook, M., & Tyers, M. (2007). Size control goes global. *Current Opinion in Biotechnology*, *18*(4), 341–50. https://doi.org/10.1016/j.copbio.2007.07.006

Coordinators, N. R. (2014). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *42*(Database issue), D7-17. https://doi.org/10.1093/nar/gkt1146

Cotton, J. A. (2005). Analytical methods for detecting paralogy in molecular datasets. *Methods in Enzymology*, *395*, 700–724. https://doi.org/10.1016/S0076-6879(05)95036-2

Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., ... D'Eustachio, P. (2014). The Reactome pathway knowledgebase. *Nucleic Acids Research*, *42*(Database issue), D472-7. https://doi.org/10.1093/nar/gkt1102

Cross, F. R., Buchler, N. E., & Skotheim, J. M. (2011). Evolution of networks and sequences in eukaryotic cell cycle control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1584), 3532–3544. https://doi.org/10.1098/rstb.2011.0078

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., ... Flicek, P. (2014). Ensembl 2015. *Nucleic Acids Research*, 1–8. https://doi.org/10.1093/nar/gku1010

Czesnick, H., & Lenhard, M. (2015). Size Control in Plants—Lessons from Leaves and Flowers. *Cold Spring Harbor Perspectives in Biology*, *7*(8), a019190. https://doi.org/10.1101/cshperspect.a019190

Davis, M. M., Boniface, J. J., Reich, Z., Lyons, D., Hampl, J., Arden, B., & Chien, Y. (1998). LIGAND RECOGNITION BY αβ T CELL RECEPTORS. *Annual Review of Immunology*, *16*(1), 523–544. https://doi.org/10.1146/annurev.immunol.16.1.523

Day, S. J., & Lawrence, P. A. (2000). Measuring dimensions: the regulation of size and shape. *Development (Cambridge, England)*, *127*(14), 2977–87. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10862736

Dayhoff, M. O. (1969). Computer analysis of protein evolution. *Scientific American*, *221*(1), 86–95. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5789703

Dayhoff, M. O., & Ledley, R. S. (1962). Comprotein: A computer program to aid primary protein structure determination. *Proc. Fall Joint Comp. Conf.*, *22*, 262–274.

de Bruin, R. A. M., McDonald, W. H., Kalashnikova, T. I., Yates, J., & Wittenberg, C. (2004). Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. *Cell*, *117*(7), 887–98. https://doi.org/10.1016/j.cell.2004.05.025

De Paiva, C. S., Pflugfelder, S. C., & Li, D.-Q. (2006). Cell size correlates with phenotype and proliferative capacity in human corneal epithelial cells. *Stem Cells (Dayton, Ohio)*, *24*(2), 368–75. https://doi.org/10.1634/stemcells.2005-0148

De Veylder, L. (2001). Functional Analysis of Cyclin-Dependent Kinase Inhibitors of Arabidopsis. *The Plant Cell Online*, *13*(7), 1653–1668. https://doi.org/10.1105/tpc.13.7.1653

Degtyarenko, K., De matos, P., Ennis, M., Hastings, J., Zbinden, M., Mcnaught, A., ... Ashburner, M. (2008). ChEBI: A database and ontology for chemical entities of biological interest. *Nucleic Acids Research*, *36*(SUPPL. 1), 344–350. https://doi.org/10.1093/nar/gkm791

Dessimoz, C., & Walker, J. M. (2017). *The Gene Ontology Handbook*. (C. Dessimoz & N. Škunca, Eds.) (Vol. 1446). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-3743-1

Dietze, H., Berardini, T. Z., Foulger, R. E., Hill, D. P., Lomax, J., Osumi-Sutherland, D., ... Mungall, C. J. (2014). TermGenie - a web-application for pattern-based ontology class generation. *Journal of Biomedical Semantics*, *5*(1), 1–13. https://doi.org/10.1186/2041-1480-5-48

Diniz, W. J. S., & Canduri, F. (2017). Bioinformatics: An overview and its applications. *Genetics and Molecular Research*, *16*(1), 1–21. https://doi.org/10.4238/gmr16019645

Dixit, R. (2013). Plant cytoskeleton: Della connects gibberellins to microtubules. *Current Biology*, *23*(11), R479–R481. https://doi.org/10.1016/j.cub.2013.04.037

Dolznig, H., Grebien, F., Sauer, T., Beug, H., & Müllner, E. W. (2004). Evidence for a size-sensing mechanism in animal cells. *Nature Cell Biology*, *6*(9), 899–905. https://doi.org/10.1038/ncb1166

Dong, J., Feldmann, G., Huang, J., Wu, S., Zhang, N., Comerford, S. A., ... Pan, D. (2007). Elucidation of a universal size-control mechanism in Drosophila and mammals. *Cell*, *130*(6), 1120–33. https://doi.org/10.1016/j.cell.2007.07.019

Dong, X., Yang, B., Li, Y., Zhong, C., & Ding, J. (2009). Molecular basis of the acceleration of the GDP-GTP exchange of human Ras homolog enriched in brain by human translationally controlled tumor protein. *Journal of Biological Chemistry*, *284*(35), 23754–23764. https://doi.org/10.1074/jbc.M109.012823

Donnelly, P. M., Bonetta, D., Tsukaya, H., Dengler, R. E., & Dengler, N. G. (1999). Cell cycling and cell enlargement in developing leaves of Arabidopsis. *Developmental Biology*, *215*(2), 407–19. https://doi.org/10.1006/dbio.1999.9443

Doolittle, R. F. (1997). Some reflections on the early days of sequence searching. *Journal of Molecular Medicine (Berlin, Germany)*, *75*(4), 239–41. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9151208

Doonan, J. H., & Sablowski, R. (2010). Walls around tumours - why plants do not develop cancer. *Nature Reviews. Cancer*, *10*(11), 794–802. https://doi.org/10.1038/nrc2942

Dornan, D., & Hupp, T. R. (2001). Inhibition of p53-dependent transcription by BOX-I phospo-peptide mimetics that bind to p300. *EMBO Reports*, *2*(2), 139–144. https://doi.org/10.1093/embo-reports/kve025

Downey, M., Houlsworth, R., Maringele, L., Rollie, A., Brehme, M., Galicia, S., ... Durocher, D. (2006). A Genome-Wide Screen Identifies the Evolutionarily Conserved KEOPS Complex as a Telomere Regulator. *Cell*, *124*(6), 1155–1168. https://doi.org/10.1016/j.cell.2005.12.044

Doyon, Y., Selleck, W., Lane, W. S., Tan, S., & Côté, J. (2004). Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast

to humans. *Molecular and Cellular Biology*, *24*(5), 1884–96. https://doi.org/10.1128/MCB.24.5.1884

Dudderidge, T. J., Kelly, J. D., Wollenschlaeger, A., Okoturo, O., Prevost, T., Robson, W., ... Stoeber, K. (2010). Diagnosis of prostate cancer by detection of minichromosome maintenance 5 protein in urine sediments. *British Journal of Cancer*, *103*(5), 701–707. https://doi.org/10.1038/sj.bjc.6605785

Dupont, S., Morsut, L., Aragona, M., Enzo, E., Giulitti, S., Cordenonsi, M., ... Piccolo, S. (2011). Role of YAP/TAZ in mechanotransduction. *Nature*, *474*(7350), 179–184. https://doi.org/10.1038/nature10137

Dustin, M. L. (2001). Role of adhesion molecules in activation signaling in T lymphocytes. *Journal of Clinical Immunology*, *21*(4), 258–63. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11506195

Eck, R. V., & Dayhoff, M. O. (1966). *Atlas of Protein Sequence and Structure*. Silver Spring, Maryland: National Biomedical Research Foundation.

Edgar, B. A. (2006). How flies get their size: genetics meets physiology. *Nature Reviews. Genetics*, *7*(12), 907–16. https://doi.org/10.1038/nrg1989

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology and Evolution*, *29*(1), 51–63. https://doi.org/10.1016/j.tree.2013.09.008

Ellis, S. A., & Ballingall, K. T. (1999). Cattle MHC: evolution in action? *Immunological Reviews*, *167*, 159–68. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10319258

Evans, T., Rosenthal, E. T., Youngblom, J., Distel, D., & Hunt, T. (1983). Cyclin: A protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. *Cell*, *33*(2), 389–396. https://doi.org/10.1016/0092-8674(83)90420-8

Facchetti, G., Chang, F., & Howard, M. (2017). Controlling cell size through sizer mechanisms. *Current Opinion in Systems Biology*, *5*, 86–92. https://doi.org/10.1016/j.coisb.2017.08.010

Facchin, S., Ruzzene, M., Peggion, C., Sartori, G., Carignani, G., Marin, O., ... Pinna, L. A. (2007). Phosphorylation and activation of the atypical kinase p53-related protein kinase (PRPK) by Akt/PKB. *Cellular and Molecular Life Sciences : CMLS*, *64*(19–20), 2680–9. https://doi.org/10.1007/s00018-007-7179-7

Famili, F., Wiekmeijer, A.-S., & Staal, F. J. (2017). The development of T cells from stem cells in mice and humans. *Future Science OA*, *3*(3), FSO186. https://doi.org/10.4155/fsoa-2016-0095

Fang, G., Bhardwaj, N., Robilotto, R., & Gerstein, M. B. (2010). Getting started in gene orthology and functional analysis. *PLoS Computational Biology*, *6*(3). https://doi.org/10.1371/journal.pcbi.1000703

Fankhauser, G. (1945). Maintenance of normal structure in heteroploid salamander larvae, through compensation of changes in cell size by adjustment of cell number and cell shape. *Journal of Experimental Zoology*, *100*(3), 445–455. https://doi.org/10.1002/jez.1401000310

Fantes, P., & Nurse, P. (1977). Control of cell size at division in fission yeast by a growth-modulated size control over nuclear division. *Experimental Cell Research*, *107*(2), 377–86. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/872891

Farmer, E., Farmer, E., Mousavi, S., & Lenglet, A. (2013). Leaf numbering for

experiments on long distance signalling in Arabidopsis. *Protocol Exchange*. https://doi.org/10.1038/protex.2013.071

Feng, Y., Liu, D., Yao, H., & Wang, J. (2007). Solution structure and mapping of a very weak calcium-binding site of human translationally controlled tumor protein by NMR. *Archives of Biochemistry and Biophysics*, *467*(1), 48–57. https://doi.org/10.1016/j.abb.2007.08.021

Ferrezuelo, F., Colomina, N., Palmisano, A., Garí, E., Gallego, C., Csikász-Nagy, A., & Aldea, M. (2012). The critical size is set at a single-cell level by growth rate to attain homeostasis and adaptation. *Nature Communications*, *3*(1012), 1–11. https://doi.org/10.1038/ncomms2015

Finetti, F., Paccani, S. R., Riparbelli, M. G., Giacomello, E., Perinetti, G., Pazour, G. J., ... Baldari, C. T. (2009). Intraflagellar transport is required for polarized recycling of the TCR/CD3 complex to the immune synapse. *Nature Cell Biology*, *11*(11), 1332–9. https://doi.org/10.1038/ncb1977

Fingar, D. C., Salama, S., Tsou, C., Harlow, E., & Blenis, J. (2002). Mammalian cell size is controlled by mTOR and its downstream targets S6K1 and 4EBP1/eIF4E. *Genes and Development*, *16*(12), 1472–1487. https://doi.org/10.1101/gad.995802

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., ... Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Research*, *42*(D1), D222–D230. https://doi.org/10.1093/nar/gkt1223

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., ... Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–D285. https://doi.org/10.1093/nar/gkv1344

Finn, R. D., Tate, J., Mistry, J., Coggill, P. C., Sammut, S. J., Hotz, H.-R., ... Bateman, A. (2008). The Pfam protein families database. *Nucleic Acids Research*, *36*(Database issue), D281-8. https://doi.org/10.1093/nar/gkm960

Fitch, W. M. (1970). Distinguishing Homologous from Analogous Proteins. *Systematic Zoology*, *19*(2), 99. https://doi.org/10.2307/2412448

Fitch, W. M. (2000). Homology - a personal view on some of the problems. *Trends in Genetics*, *16*(5), 227–231. https://doi.org/10.1016/S0168-9525(00)02005-9

Fleischer, T. C., Weaver, C. M., McAfee, K. J., Jennings, J. L., & Link, A. J. (2006). Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. *Genes and Development*, *20*(10), 1294–1307. https://doi.org/10.1101/gad.1422006

Forsburg, S. L. (2004). Eukaryotic MCM Proteins: Beyond Replication Initiation. *Microbiology and Molecular Biology Reviews*, *68*(1), 109–131. https://doi.org/10.1128/MMBR.68.1.109-131.2004

Fox, C. J. C., Hammerman, P. P. S., & Thompson, C. B. C. (2005). Fuel feeds function: Energy metabolism and the T-cell response. *Nature Reviews Immunology*, *5*(11), 844–852. https://doi.org/10.1038/nri1710

Freeman, G. J., Gribben, J. G., Boussiotis, V. A., Ng, J. W., Restivo, V. A., Lombard, L. A., ... Nadler, L. M. (1993). Cloning of B7-2: a CTLA-4 counter-receptor that costimulates human T cell proliferation. *Science (New York, N.Y.)*, *262*(5135), 909–11. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7694363

Fu, X., He, F., Li, Y., Shahveranov, A., & Hutchins, A. P. (2017). Genomic and molecular control of cell type and cell type conversions. *Cell Regeneration*, *6*, 1–7.

https://doi.org/10.1016/j.cr.2017.09.001

Fujioka, Y., Taira, T., Maeda, Y., Tanaka, S., Nishihara, H., Iguchi-Ariga, S. M. M., ... Ariga, H. (2001). NM-1, a c-Myc-binding Protein, Is a Candidate for a Tumor Suppressor in Leukemia/Lymphoma and Tongue Cancer. *Journal of Biological Chemistry*, *276*(48), 45137–45144. https://doi.org/10.1074/jbc.M106127200

Fujiwara, M., Sengupta, P., & McIntire, S. L. (2002). Regulation of body size and behavioral state of C. elegans by sensory perception and the egl-4 cGMP-dependent protein kinase. *Neuron*, *36*(6), 1091–1102. https://doi.org/10.1016/S0896-6273(02)01093-0

Gachet, Y., Tournier, S., Lee, M., Lazaris-karatzas, A., & Poulton, T. (1999). The growth-related , translationally controlled protein P23 has properties of a tubulin binding protein and associates transiently with microtubules during the cell cycle. *Journal of Cell Science*, *1271*(Pt 8), 1257–1271.

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2015). Expanded Microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Research*, *43*(D1), D261–D269. https://doi.org/10.1093/nar/gku1223

Garcia, D. (2003). Arabidopsis haiku Mutants Reveal New Controls of Seed Size by Endosperm. *Plant Physiology*, *131*(4), 1661–1670. https://doi.org/10.1104/pp.102.018762

Gasser, C. S., Broadhvest, J., & Hauser, B. A. (1998). Genetic Analysis of Ovule Development. *Annual Review of Plant Physiology and Plant Molecular Biology*, *49*(1), 1–24. https://doi.org/10.1146/annurev.arplant.49.1.1

Gaudet, P., Livstone, M. S., Lewis, S. E., & Thomas, P. D. (2011). Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Briefings in Bioinformatics*, *12*(5), 449–462. https://doi.org/10.1093/bib/bbr042

Geissler, S., Siegers, K., & Schiebel, E. (1998). A novel protein complex promoting formation of functional alpha- and gamma-tubulin. *The EMBO Journal*, *17*(4), 952–66. https://doi.org/10.1093/emboj/17.4.952

Gene Ontology Consortium. (2015). Gene Ontology Consortium: going forward. *Nucleic Acids Research*, *43*(D1), D1049–D1056. https://doi.org/10.1093/nar/gku1179

Genevet, A., & Tapon, N. (2011). The Hippo pathway and apico–basal cell polarity. *Biochemical Journal*, *436*(2), 213–224. https://doi.org/10.1042/BJ20110217

Gerriets, V. A., & Rathmell, J. C. (2012). Metabolic pathways in T cell fate and function. *Trends in Immunology*, *33*(4), 168–172. https://doi.org/10.1016/j.it.2012.01.010

Gkoutos, G. V, Green, E. C. J., Mallon, A., Hancock, J. M., & Davidson, D. (2005). Using ontologies to describe mouse phenotypes. *Genome Biology*, *6*(1), R8. https://doi.org/10.1186/gb-2004-6-1-r8

Gnanasekar, M., Dakshinamoorthy, G., & Ramaswamy, K. (2009). Translationally controlled tumor protein is a novel heat shock protein with chaperone-like activity. *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/j.bbrc.2009.06.028

Goberdhan, D. C. I., & Wilson, C. (2003). The functions of insulin signaling: size isn't everything, even in Drosophila. *Differentiation*, *71*(7), 375–397. https://doi.org/10.1046/j.1432-0436.2003.7107001.x

Goldstein, B., & Hird, S. N. (1996). Specification of the anteroposterior axis in Caenorhabditis elegans. *Development (Cambridge, England)*, *122*(5), 1467–1474.

Gonzalez, I. L., & Schmickel, R. D. (1986). The human 18S ribosomal RNA gene: evolution and stability. *American Journal of Human Genetics*, *38*(4), 419–27. https://doi.org/citeulike-article-id:605092

Gonzalez, S., & Rallis, C. (2017). The TOR Signaling Pathway in Spatial and Temporal Control of Cell Size and Growth. *Frontiers in Cell and Developmental Biology*, *5*(June), 1–6. https://doi.org/10.3389/fcell.2017.00061

Goodman, S. R. (2008). *Medical Cell Biology*. (S. R. Goodman, Ed.) (3rd ed.). Elsevier Inc. https://doi.org/10.1016/B978-0-12-370458-0.50015-3

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., ... Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*(10), 3420–3435. https://doi.org/10.1093/nar/gkn176

Graml, V., Studera, X., Lawson, J. L. D., Chessel, A., Geymonat, M., Bortfeld-Miller, M., ... Carazo-Salas, R. E. (2014). A Genomic Multiprocess Survey of Machineries that Control and Link Cell Shape, Microtubule Organization, and Cell-Cycle Progression. *Developmental Cell*, *31*(2), 227–239. https://doi.org/10.1016/j.devcel.2014.09.005

Gross, B., Gaestel, M., Böhm, H., & Bielka, H. (1989). cDNA sequence coding for a translationally controlled human tumor protein. *Nucleic Acids Research*, *17*(20), 8367. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2813067

Gross, J. A., St John, T., & Allison, J. P. (1990). The murine homologue of the T lymphocyte antigen CD28. Molecular cloning and cell surface expression. *Journal of Immunology (Baltimore, Md.: 1950)*, *144*(8), 3201–10. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2157764

Grossniklaus, U., Vielle-Calzada, J.-P., Hoeppner, M. A., & Gagliano, W. B. (1998). Maternal Control of Embryogenesis by MEDEA, a Polycomb Group Gene in Arabidopsis thaliana. *Science*, *280*(February), 446–450.

Gstaiger, M., Luke, B., Hess, D., Oakeley, E. J., Wirbelauer, C., Blondel, M., ... Krek, W. (2003). Control of Nutrient-Sensitive Transcription Programs by the Unconventional Prefoldin URI. *Science*, *302*(5648), 1208–1212. https://doi.org/10.1126/science.1088401

Gu, Y., Deng, Z., Paredez, A. R., DeBolt, S., Wang, Z.-Y., & Somerville, C. (2008). Prefoldin 6 is required for normal microtubule dynamics and organization in Arabidopsis. *Proceedings of the National Academy of Sciences*, *105*(46), 18064–18069. https://doi.org/10.1073/pnas.0808652105

Guertin, D. A., & Sabatini, D. M. (2006). Cell size control. In *Encyclopedia of Life Sciences*. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1038/npg.els.0003359

Gupta, M., Yoshioka, H., Ohnishi, K., Mizumoto, H., Hikichi, Y., & Kiba, A. (2013). A translationally controlled tumor protein negatively regulates the hypersensitive response in nicotiana benthamiana. *Plant and Cell Physiology*, *54*(8), 1403–1414. https://doi.org/10.1093/pcp/pct090

Gutiérrez-Galeano, D. F., Toscano-Morales, R., Calderón-Pérez, B., Xoconostle-Cázares, B., & Ruiz-Medrano, R. (2014). Structural divergence of plant TCTPs. *Frontiers in Plant Science*, *5*, 361. https://doi.org/10.3389/fpls.2014.00361

Haber, A. H. (1962). Non-essentiality of concurrent cell divisions for degree of polarization of leaf growth. I. Studies with radiation-induced mitotic inhibition. *American Journal of Botany*. https://doi.org/10.2307/2439715

Hagen, J. B. (2000). The origins of bioinformatics. *Nature Reviews Genetics*, *1*(3), 231–236. https://doi.org/10.1038/35042090

Haghighat, N. G., & Ruben, L. (1992). Purification of novel calcium binding proteins from Trypanosoma brucei: properties of 22-, 24- and 38-kilodalton proteins. *Molecular and Biochemical Parasitology*, *51*(1), 99–110. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1565142

Halder, G., & Johnson, R. L. (2011). Hippo signaling: growth control and beyond. *Development*, *138*(1), 9–22. https://doi.org/10.1242/dev.045500

Hamilton, S. E., & Jameson, S. C. (2012). CD8 T cell quiescence revisited. *Trends in Immunology*, *33*(5), 224–230. https://doi.org/10.1016/j.it.2012.01.007

Hammerling, J. (1963). Nucleo-cytoplasmic interactions in Acetabularia and other cells. *Annu Rev Plant Physiol 1963, 14*, 65–92.

Hartl, F. U. (1996). Molecular chaperones in cellular protein folding. *Nature*. https://doi.org/10.1038/381571a0

Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature*, *475*(7356), 324–332. https://doi.org/10.1038/nature10317

Hartl, F. U., & Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science (New York, N.Y.)*, *295*(5561), 1852–8. https://doi.org/10.1126/science.1068408

Hartwell, L. H., Culotti, J., & Reid, B. (1970). Genetic control of the cell-division cycle in yeast. I. Detection of mutants. *Proceedings of the National Academy of Sciences of the United States of America*, *66*(2), 352–9. https://doi.org/10.1073/pnas.66.2.352

Harvey, K. F., Zhang, X., & Thomas, D. M. (2013). The Hippo pathway and human cancer. *Nature Reviews Cancer*, *13*(4), 246–257. https://doi.org/10.1038/nrc3458

Haupt, Y., Maya, R., Kazaz, A., & Oren, M. (1997). Mdm2 promotes the rapid degradation of p53. *Nature*, *387*(6630), 296–299. https://doi.org/10.1038/387296a0

Hawkins, N., & Garriga, G. (1998). Asymmetric cell division: from A to Z. *Genes & Development*, *12*(23), 3625–3638. https://doi.org/10.1101/gad.12.23.3625

Hay, N., & Sonenberg, N. (2004). Upstream and downstream of mTOR. *Genes & Development*, *18*(16), 1926–45. https://doi.org/10.1101/gad.1212704

Hayflick, L. (1965). The Limited in vitro lifetime of human diploid cell strains. *Experimental Cell Research*, *37*, 614–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14315085

Hayles, J., Wood, V., Jeffery, L., Hoe, K., Kim, D., Park, H., ... Nurse, P. (2013). A genome-wide resource of cell cycle and cell shape genes of fission yeast. *Open Biology*, *3*(5), 130053. https://doi.org/http://dx.doi.org/10.1098/rsob.130053

He, S., Huang, Y., Wang, Y., Tang, J., Song, Y., Yu, X., ... Xu, X. (2015). Histamine-releasing factor/translationally controlled tumor protein plays a role in induced cell adhesion, apoptosis resistance and chemoresistance in non-Hodgkin lymphomas. *Leukemia and Lymphoma*, *56*(7), 2153–2161. https://doi.org/10.3109/10428194.2014.981173

Heidelberg, J. F., Paulsen, I. T., Nelson, K. E., Gaidos, E. J., Nelson, W. C., Read, T.

D., ... Fraser, C. M. (2002). Genome sequence of the dissimilatory metal ion–reducing bacterium Shewanella oneidensis. *Nature Biotechnology*, *20*(11), 1118–1123. https://doi.org/10.1038/nbt749

Heitman, J., Movva, N. R., & Hall, M. N. (1991). Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. *Science*, *253*(5022), 905–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1715094

Hemerly, A., Engler, J. de A., Bergounioux, C., Van Montagu, M., Engler, G., Inzé, D., & Ferreira, P. (1995). Dominant negative mutants of the Cdc2 kinase uncouple cell division from iterative plant development. *The EMBO Journal*, *14*(16), 3925–36. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC394471/pdf/emboj00040-0083.pdf

Hengstschläger, M., Braun, K., Soucek, T., Miloloza, A., & Hengstschläger-Ottnad, E. (1999). Cyclin-dependent kinases at the G1-S transition of the mammalian cell cycle. *Mutation Research*, *436*(1), 1–9. https://doi.org/https://doi.org/10.1016/S1383-5742(98)00022-2

Herridge, R. P., Day, R. C., Baldwin, S., & Macknight, R. C. (2011). Rapid analysis of seed size in Arabidopsis for mutant and QTL discovery. *Plant Methods*, *7*(1), 3. https://doi.org/10.1186/1746-4811-7-3

Herridge, R. P., Day, R. C., & Macknight, R. C. (2014). The role of the MCM2-7 helicase complex during Arabidopsis seed development. *Plant Molecular Biology*, *86*(1–2), 69–84. https://doi.org/10.1007/s11103-014-0213-x

Hill, J. E., & Hemmingsen, S. M. (2001). Arabidopsis thaliana type I and II chaperonins. *Cell Stress & Chaperones*, *6*(3), 190–200. https://doi.org/10.1379/1466-1268(2001)006<0190:ATTIAI>2.0.CO;2

Hinojosa-Moya, J. J., Xoconostle-Cázares, B., Toscano-Morales, R., Ramírez-Ortega, F., Cabrera-Ponce, J. L., & Ruiz-Medrano, R. (2013). Characterization of the pumpkin Translationally-Controlled Tumor Protein CmTCTP. *Plant Signaling & Behavior*, *8*(12), e26477. https://doi.org/10.4161/psb.26477

Hinojosa-Moya, J., Xoconostle-Cázares, B., Piedra-Ibarra, E., Méndez-Tenorio, A., Lucas, W. J., & Ruiz-Medrano, R. (2008). Phylogenetic and structural analysis of translationally controlled tumor proteins. *Journal of Molecular Evolution*, *66*(5), 472–483. https://doi.org/10.1007/s00239-008-9099-z

Hirao, A., Kong, Y. Y., Matsuoka, S., Wakeham, A., Ruland, J., Yoshida, H., ... Mak, T. W. (2000). DNA damage-induced activation of p53 by the checkpoint kinase Chk2. *Science (New York, N.Y.)*, *287*(5459), 1824–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10710310

Hirschman, J. E., Balakrishnan, R., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., ... Cherry, J. M. (2006). Genome Snapshot: a new resource at the Saccharomyces Genome Database (SGD) presenting an overview of the Saccharomyces cerevisiae genome. *Nucleic Acids Research*, *34*(Database issue), D442-5. https://doi.org/10.1093/nar/gkj117

Hochegger, H., Takeda, S., & Hunt, T. (2008). Cyclin-dependent kinases and cell-cycle transitions: Does one fit all? *Nature Reviews Molecular Cell Biology*, *9*(11), 910–916. https://doi.org/10.1038/nrm2510

Hoffman, C. S., Wood, V., & Fantes, P. A. (2015). An ancient yeast for young geneticists: A primer on the Schizosaccharomyces pombe model system. *Genetics*,

201(2), 403–423. https://doi.org/10.1534/genetics.115.181503

Holland, L., Downey, M., Song, X., Gauthier, L., Bell-Rogers, P., & Yankulov, K. (2002). Distinct parts of minichromosome maintenance protein 2 associate with histone H3/H4 and RNA polymerase II holoenzyme. *European Journal of Biochemistry*, *269*(21), 5192–5202. https://doi.org/10.1046/j.1432-1033.2002.03224.x

Honda, R., Tanaka, H., & Yasuda, H. (1997). Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. *FEBS Letters*, *420*(1), 25–27. https://doi.org/10.1016/S0014-5793(97)01480-4

Hong, S.-T., & Choi, K.-W. (2013). TCTP directly regulates ATM activity to control genome stability and organ development in Drosophila melanogaster. *Nature Communications*, *4*, 1–14. https://doi.org/10.1038/ncomms3986

Hooke, R. (1665). *Micrographia*. *Council of the Royal Society of London for Improving Natural Knowledge London United Kingdom*. https://doi.org/10.5962/bhl.title.904

Horiguchi, G., Kim, G.-T., & Tsukaya, H. (2005). The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of Arabidopsis thaliana. *The Plant Journal : For Cell and Molecular Biology*, *43*(1), 68–78. https://doi.org/10.1111/j.1365-313X.2005.02429.x

Horiguchi, G., & Tsukaya, H. (2011). Organ Size Regulation in Plants: Insights from Compensation. *Frontiers in Plant Science*, *2*(July), 1–6. https://doi.org/10.3389/fpls.2011.00024

Hsu, Y. C., Chern, J. J., Cai, Y., Liu, M., & Choi, K. W. (2007). Drosophila TCTP is essential for growth and proliferation through regulation of dRheb GTPase. *Nature*, *445*(7129), 785–788. https://doi.org/10.1038/nature05528

Hu, W., Feng, Z., & Levine, A. J. (2012). The Regulation of Multiple p53 Stress Responses is Mediated through MDM2. *Genes and Cancer*, *3*(3–4), 199–208. https://doi.org/10.1177/1947601912454734

Huala, E., Dickerman, A. W., Garcia-hernandez, M., Weems, D., Reiser, L., Lafond, F., ... Rhee, S. Y. (2001). The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Research*, *29*(1), 102–105.

Huang, J., Wu, S., Barrera, J., Matthews, K., & Pan, D. (2005). The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the Drosophila homolog of YAP. *Cell*, *122*(3), 421–434. https://doi.org/10.1016/j.cell.2005.06.007

Huber, M. D., & Gerace, L. (2007). The size-wise nucleus: Nuclear volume control in eukaryotes. *Journal of Cell Biology*, *179*(4), 583–584. https://doi.org/10.1083/jcb.200710156

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., ... Bork, P. (2016). EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, *44*(D1), D286–D293. https://doi.org/10.1093/nar/gkv1248

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., ... Bork, P. (2018). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(November 2018), 309–314.

https://doi.org/10.1093/nar/gky1085

Hughes, A. L. (1995). Origin and evolution of hla class-i pseudogenes. *Molecular Biology and Evolution*, *12*(December), 247–258.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuche, B. A., de Castro, E., ... Sigrist, C. J. A. (2008). The 20 years of PROSITE. *Nucleic Acids Research*, *36*(Database issue), D245-9. https://doi.org/10.1093/nar/gkm977

Hunt, L. T. (1983). Margaret O. Dayhoff 1925-1983. *DNA (Mary Ann Liebert, Inc.)*, *2*(2), 97–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6347589

Hunter, M., Scourfield, E. J., Emmott, E., & Graham, S. C. (2017). VPS18 recruits VPS41 to the human HOPS complex via a RING-RING interaction. *Biochemical Journal*, *0*, BCJ20170588. https://doi.org/10.1042/BCJ20170588

Huntley, R. P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M. J., & O'Donovan, C. (2015). The GOA database: Gene Ontology annotation updates for 2015. *Nucleic Acids Research*, *43*(D1), D1057–D1063. https://doi.org/10.1093/nar/gku1113

Huntley, R. P., Sitnikov, D., Orlic-Milacic, M., Balakrishnan, R., D'Eustachio, P., Gillespie, M. E., ... Lovering, R. C. (2016). Guidelines for the functional annotation of microRNAs using the Gene Ontology. *RNA (New York, N.Y.)*, *22*(5), 667–76. https://doi.org/10.1261/rna.055301.115

Hutchins, J. R. A., Toyoda, Y., Hegemann, B., Poser, I., Hériché, J.-K., Sykora, M. M., ... Peters, J.-M. (2010). Systematic analysis of human protein complexes identifies chromosome segregation proteins. *Science (New York, N.Y.)*, *328*(5978), 593–9. https://doi.org/10.1126/science.1181348

Ibar, C., Cataldo, V. F., Vasquez-Doorman, C., Olguin, P., & Glavic, A. (2013). Drosophila p53-related protein kinase is required for PI3K/TOR pathway-dependent growth. *Development*, *140*(6), 1282–1291. https://doi.org/10.1242/dev.086918

Isenbarger, T. A., Carr, C. E., Johnson, S. S., Finney, M., Church, G. M., Gilbert, W., ... Ruvkun, G. (2008). The most conserved genome segments for life detection on earth and other planets. *Origins of Life and Evolution of Biospheres*, *38*(6), 517–533. https://doi.org/10.1007/s11084-008-9148-z

Jaglarz, M. K., Bazile, F., Laskowska, K., Polanski, Z., Chesnel, F., Borsuk, E., ... Kubiak, J. Z. (2012). Association of TCTP with centrosome and microtubules. *Biochemistry Research International*, *2012*. https://doi.org/10.1155/2012/541906

Jaiswal, P., Avraham, S., Ilic, K., Kellogg, E. A., McCouch, S., Pujar, A., ... Zapata, F. (2005). Plant Ontology (PO): A controlled vocabulary of plant structures and growth stages. *Comparative and Functional Genomics*, *6*(7–8), 388–397. https://doi.org/10.1002/cfg.496

Jeffrey, P. D., Tong, L., & Pavletich, N. P. (2000). Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors. *Genes and Development*, *14*(24), 3115–3125. https://doi.org/10.1101/gad.851100

Jenkins, M. K., Chen, C. A., Jung, G., Mueller, D. L., & Schwartz, R. H. (1990). Inhibition of antigen-specific proliferation of type 1 murine T cell clones after stimulation with immobilized anti-CD3 monoclonal antibody. *Journal of Immunology (Baltimore, Md. : 1950)*, *144*(1), 16–22. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2153162

Jensen, L. J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., & Bork, P.

(2008). eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Research*, *36*(Database issue), D250-4. https://doi.org/10.1093/nar/gkm796

Jensen, R. a. (2001). Orthologs and paralogs - we need to get it right. *Genome Biology*, *2*(8), interactions1002.1-interactions1002.3. https://doi.org/10.1186/gb-2001-2-8-interactions1002

Johnson, C. S. (2002). TRANSPARENT TESTA GLABRA2, a Trichome and Seed Coat Development Gene of Arabidopsis, Encodes a WRKY Transcription Factor. *The Plant Cell Online*, *14*(6), 1359–1375. https://doi.org/10.1105/tpc.001404

Johnson, L. S., Eddy, S. R., & Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics*, *11*, 431. https://doi.org/10.1186/1471-2105-11-431

Jorgensen, P., Edgington, N. P., Schneider, B. L., Rupes, I., Tyers, M., & Futcher, B. (2007). The size of the nucleus increases as yeast cells grow. *Molecular Biology of the Cell*, *18*(9), 3523–32. https://doi.org/10.1091/mbc.E06-10-0973

Jorgensen, P., Nishikawa, J. L., Breitkreutz, B.-J., & Tyers, M. (2002). Systematic identification of pathways that couple cell growth and division in yeast. *Science*, *297*(5580), 395–400. https://doi.org/10.1126/science.1070850

Jorgensen, P., Rupes, I., Sharom, J. R., Schneper, L., Broach, J. R., & Tyers, M. (2004). A dynamic transcriptional network communicates growth potential to ribosome synthesis and critical cell size. *Genes & Development*, *18*(20), 2491–505. https://doi.org/10.1101/gad.1228804

Jorgensen, P., & Tyers, M. (2004). How cells coordinate growth and division. *Current Biology : CB*, *14*(23), R1014-27. https://doi.org/10.1016/j.cub.2004.11.027

Jung, J., Kim, H. Y., Kim, M., Sohn, K., Kim, M., & Lee, K. (2011). Translationally controlled tumor protein induces human breast epithelial cell transformation through the activation of Src. *Oncogene*, *30*(19), 2264–2274. https://doi.org/10.1038/onc.2010.604

Kalra, S., Joshi, G., Munshi, A., & Kumar, R. (2017). *Structural insights of cyclin dependent kinases: Implications in design of selective inhibitors*. *European Journal of Medicinal Chemistry* (Vol. 142). Elsevier Masson SAS. https://doi.org/10.1016/j.ejmech.2017.08.071

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Research*, *42*(D1), 199–205. https://doi.org/10.1093/nar/gkt1076

Karman, J., Jiang, J.-L., Gumlaw, N., Zhao, H., Campos-Rivera, J., Sancho, J., ... Zhu, Y. (2012). Ligation of Cytotoxic T Lymphocyte Antigen-4 to T Cell Receptor Inhibits T Cell Activation and Directs Differentiation into Foxp3 $^{+}$ Regulatory T Cells. *Journal of Biological Chemistry*, *287*(14), 11098–11107. https://doi.org/10.1074/jbc.M111.283705

Kataoka, N., Bachorik, J. L., & Dreyfuss, G. (1999). Transportin-SR, a nuclear import receptor for SR proteins. *Journal of Cell Biology*, *145*(6), 1145–1152. https://doi.org/10.1083/jcb.145.6.1145

Kim, D. K., Nam, B. Y., Li, J. J., Park, J. T., Lee, S. H., Kim, D. H., ... Kang, S. W. (2012). Translationally controlled tumour protein is associated with podocyte hypertrophy in a mouse model of type 1 diabetes. *Diabetologia*, *55*(4), 1205–1217.

https://doi.org/10.1007/s00125-012-2467-7

Kim, S., Wong, P., & Coulombe, P. A. (2006). A keratin cytoskeletal protein regulates protein synthesis and epithelial cell growth. *Nature*, *441*(7091), 362–365. https://doi.org/10.1038/nature04659

Kinsella, R. J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., ... Flicek, P. (2011). Ensembl BioMarts: A hub for data retrieval across taxonomic space. *Database*, *2011*, 1–9. https://doi.org/10.1093/database/bar030

Kisseleva-Romanova, E., Lopreiato, R., Baudin-Baillieu, A., Rousselle, J. C., Ilan, L., Hofmann, K., ... Libri, D. (2006). Yeast homolog of a cancer-testis antigen defines a new transcription complex. *EMBO Journal*, *25*(15), 3576–3585. https://doi.org/10.1038/sj.emboj.7601235

Kiyokawa, H., Kineman, R. D., Manova-Todorova, K. O., Soares, V. C., Hoffman, E. S., Ono, M., ... Beach, D. (1996). Enhanced growth of mice lacking the cyclin-dependent kinase inhibitor function of p27(Kip1). *Cell*, *85*(5), 721–32. https://doi.org/10.1016/s0092-8674(00)81238-6

Knoll, A. H. (1992). The early evolution of eukaryotes: a geological perspective. *Science*, *256*(5057), 622–627. https://doi.org/10.1126/science.1585174

Koide, Y., Kiyota, T., Tonganunt, M., Pinkaew, D., Liu, Z., Kato, Y., ... Fujise, K. (2009). Embryonic lethality of fortilin-null mutant mice by BMP-pathway overactivation. *Biochimica et Biophysica Acta - General Subjects*, *1790*(5), 326–338. https://doi.org/10.1016/j.bbagen.2009.01.012

Koonin, E. V, Mushegian, A. R., & Bork, P. (1996). Non-orthologous gene displacement. *Trends in Genetics : TIG*, *12*(9), 334–6.

Kordasti, S., Costantini, B., Seidl, T., Perez Abellan, P., Martinez Llordella, M., McLornan, D., ... Mufti, G. J. (2016). Deep-phenotyping of Tregs identifies an immune signature for idiopathic aplastic anemia and predicts response to treatment. *Blood*. https://doi.org/10.1182/blood-2016-03-703702

Krishnan, L., Matreyek, K. A., Oztop, I., Lee, K., Tipper, C. H., Li, X., ... Engelman, A. (2010). The Requirement for Cellular Transportin 3 (TNPO3 or TRN-SR2) during Infection Maps to Human Immunodeficiency Virus Type 1 Capsid and Not Integrase. *Journal of Virology*, *84*(1), 397–406. https://doi.org/10.1128/JVI.01899-09

Krogsgaard, M., & Davis, M. M. (2005). How T cells "see" antigen. *Nature Immunology*, *6*(3), 239–245. https://doi.org/10.1038/ni1173

Kuhn, M., Szklarczyk, D., Pletscher-Frankild, S., Blicher, T. H., Von Mering, C., Jensen, L. J., & Bork, P. (2014). STITCH 4: Integration of protein-chemical interactions with user data. *Nucleic Acids Research*, *42*(D1), 401–407. https://doi.org/10.1093/nar/gkt1207

Kuhns, M. S., Epshteyn, V., Sobel, R. A., & Allison, J. P. (2000). Cytotoxic T lymphocyte antigen-4 (CTLA-4) regulates the size, reactivity, and function of a primed pool of CD4+ T cells. *Proceedings of the National Academy of Sciences*, *97*(23), 12711–12716. https://doi.org/10.1073/pnas.220423597

Kunz, J., Henriquez, R., Schneider, U., Deuter-Reinhard, M., Movva, N. R., & Hall, M. N. (1993). Target of rapamycin in yeast, TOR2, is an essential phosphatidylinositol kinase homolog required for G1 progression. *Cell*, *73*(3), 585–596. https://doi.org/10.1016/0092-8674(93)90144-F

Kuse, R., Schuster, S., Schübbe, H., Dix, S., & Hausmann, K. (1985). Blood

lymphocyte volumes and diameters in patients with chronic lymphocytic leukemia and normal controls. *Blut*, *50*(4), 243–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3857080

Lai, M.-C., Lin, R.-I., & Tarn, W.-Y. (2001). Transportin-SR2 mediates nuclear import of phosphorylated SR proteins. *Proceedings of the National Academy of Sciences*, *98*(18), 10154–10159. https://doi.org/10.1073/pnas.181354098

Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M. R., Ahn, T.-H., ... Ussery, D. W. (2015). Insights from 20 years of bacterial genome sequencing. *Functional & Integrative Genomics*, *15*(2), 141–161. https://doi.org/10.1007/s10142-015-0433-4

Lau, O. S., & Deng, X. W. (2010). Plant hormone signaling lightens up: integrators of light and hormones. *Current Opinion in Plant Biology*. https://doi.org/10.1016/j.pbi.2010.07.001

Lea, N. C., Buggins, A. G. S., Orr, S. J., Mufti, G. J., & Thomas, N. S. B. (2003). High efficiency protein transduction of quiescent and proliferating primary hematopoietic cells. *Journal of Biochemical and Biophysical Methods*, *55*(3), 251–258. https://doi.org/10.1016/S0165-022X(03)00077-0

Lea, N. C., Orr, S. J., Stoeber, K., Williams, G. H., Lam, E. W.-F., Ibrahim, M. a a, ... Thomas, N. S. B. (2003). Commitment point during G0-->G1 that controls entry into the cell cycle. *Molecular and Cellular Biology*, *23*(7), 2351–61. https://doi.org/10.1128/MCB.23.7.2351

Lee, J. H., Kirkham, J. C., McCormack, M. C., Medina, M. A., Nicholls, A. M., Randolph, M. A., & Austen, W. G. (2012). A novel approach to adipocyte analysis. *Plastic and Reconstructive Surgery*, *129*(2), 380–387. https://doi.org/10.1097/PRS.0b013e31823aea29

Lee, K. E., Ambrose, Z., Martin, T. D., Oztop, I., Mulky, A., Julias, J. G., ... KewalRamani, V. N. (2010). Flexible Use of Nuclear Import Pathways by HIV-1. *Cell Host and Microbe*, *7*(3), 221–233. https://doi.org/10.1016/j.chom.2010.02.007

Lee, M. G., & Nurse, P. (1987). Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. *Nature*, *327*(6117), 31–5. https://doi.org/10.1038/327031a0

Lee, Y., Smith, R. S., Jordan, W., King, B. L., Won, J., Valpuesta, J. M., ... Nishina, P. M. (2011). Prefoldin 5 is required for normal sensory and neuronal development in a murine model. *Journal of Biological Chemistry*, *286*(1), 726–736. https://doi.org/10.1074/jbc.M110.177352

Lees-Miller, S. P., Sakaguchi, K., Ullrich, S. J., Appella, E., & Anderson, C. W. (1992). Human DNA-activated protein kinase phosphorylates serines 15 and 37 in the amino-terminal transactivation domain of human p53. *Molecular and Cellular Biology*, *12*(11), 5041–5049. https://doi.org/10.1128/MCB.12.11.5041.Updated

Leroux, M. R., & Hartl, F. U. (2000). Protein folding: Versatility of the cytosolic chaperonin TRiC/CCT. *Current Biology*, *10*(7), 260–264. https://doi.org/10.1016/S0960-9822(00)00432-2

Levine, A. J. (1997). P53, the Cellular Gatekeeper for Growth and Division. *Cell*, *88*(3), 323–331. https://doi.org/10.1016/S0092-8674(00)81871-1

Li, D., Deng, Z., Liu, X., & Qin, B. (2013). Molecular cloning, expression profiles and characterization of a novel translationally controlled tumor protein in rubber tree

(Hevea brasiliensis). *Journal of Plant Physiology*, *170*(5), 497–504. https://doi.org/10.1016/j.jplph.2012.11.014

Li, F., Zhang, D., & Fujise, K. (2001). Characterization of Fortilin, a Novel Antiapoptotic Protein. *Journal of Biological Chemistry*, *276*(50), 47542–47549. https://doi.org/10.1074/jbc.M108954200

Li, H., Coghlan, A., Ruan, J., Coin, L. J., Hériché, J.-K., Osmotherly, L., ... Durbin, R. (2006). TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Research*, *34*, D572–D580. https://doi.org/10.1093/nar/gkj118

Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*(9), 2178–89. https://doi.org/10.1101/gr.1224503

Li, S., Chen, M., Xiong, Q., Zhang, J., Cui, Z., & Ge, F. (2016). Characterization of the Translationally Controlled Tumor Protein (TCTP) Interactome Reveals Novel Binding Partners in Human Cancer Cells. *Journal of Proteome Research*, *15*(10), 3741–3751. https://doi.org/10.1021/acs.jproteome.6b00556

Li, Y., Sedwick, C. E., Hu, J., & Altman, A. (2005). Role for Protein Kinase Cθ (PKCθ) in TCR/CD28-mediated Signaling through the Canonical but Not the Non-canonical Pathway for NF-κB Activation. *Journal of Biological Chemistry*, *280*(2), 1217–1223. https://doi.org/10.1074/jbc.M409492200

Li, Y., Zheng, L., Corke, F., Smith, C., & Bevan, M. W. (2008). Control of final seed and organ size by the *DA1* gene family in *Arabidopsis thaliana*. *Genes & Development*, *22*, 1331–1336. https://doi.org/10.1101/gad.463608.spite

Liu, H., Peng, H.-W., Cheng, Y.-S., Yuan, H. S., & Yang-Yen, H.-F. (2005). Stabilization and Enhancement of the Antiapoptotic Activity of Mcl-1 by TCTP. *Molecular and Cellular Biology*, *25*(8), 3117–3126. https://doi.org/10.1128/MCB.25.8.3117-3126.2005

Liu, S., Ginzberg, M. B., Patel, N., Hild, M., Leung, B., Li, Z., ... Kafri, R. (2018). Size uniformity of animal cells is actively maintained by a p38 MAPK-dependent regulation of G1-length. *ELife*, *7*(734), 1–26. https://doi.org/10.7554/eLife.26947

Lloyd, A. C. (2013). The Regulation of Cell Size. *Cell*, *154*(6), 1194–1205. https://doi.org/10.1016/j.cell.2013.08.053

Lloyd, J., & Meinke, D. (2012). A Comprehensive Dataset of Genes with a Loss-of-Function Mutant Phenotype in Arabidopsis. *PLANT PHYSIOLOGY*, *158*(3), 1115–1129. https://doi.org/10.1104/pp.111.192393

Lo Surdo, J., & Bauer, S. R. (2012). Quantitative Approaches to Detect Donor and Passage Differences in Adipogenic Potential and Clonogenicity in Human Bone Marrow-Derived Mesenchymal Stem Cells. *Tissue Engineering Part C: Methods*, *18*(11), 877–889. https://doi.org/10.1089/ten.tec.2011.0736

Locascio, A., Blázquez, M. A., & Alabadí, D. (2013). Dynamic regulation of cortical microtubule organization through prefoldin-DELLA interaction. *Current Biology*, *23*(9), 804–809. https://doi.org/10.1016/j.cub.2013.03.053

Loewenstein, Y., Raimondo, D., Redfern, O. C., Watson, J., Frishman, D., Linial, M., ... Tramontano, A. (2009). Protein function annotation by homology-based inference. *Genome Biology*, *10*(2), 207. https://doi.org/10.1186/gb-2009-10-2-207

López-Amorós, R., Comas, J., Carulla, C., & Vives-Rego, J. (1994). Variations in flow

cytometric forward scatter signals and cell size in batch cultures of Escherichia coli. *FEMS Microbiology Letters*, *117*(2), 225–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8181728

Loughery, J., Cox, M., Smith, L. M., & Meek, D. W. (2014). Critical role for p53-serine 15 phosphorylation in stimulating transactivation at p53-responsive promoters. *Nucleic Acids Research*, *42*(12), 7666–7680. https://doi.org/10.1093/nar/gku501

Lucas, P. J., Negishi, I., Nakayama, K., Fields, L. E., & Loh, D. Y. (1995). Naive CD28-deficient T cells can initiate but not sustain an in vitro antigen-specific immune response. *Journal of Immunology (Baltimore, Md. : 1950)*, *154*(11), 5757–68. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7751626

Lucibello, M., Adanti, S., Antelmi, E., Dezi, D., Ciafrè, S., Carcangiu, M. L., ... Pierimarchi, P. (2015). Phospho-TCTP as a therapeutic target of Dihydroartemisinin for aggressive breast cancer cells. *Oncotarget*, *6*(7), 5275–91. https://doi.org/10.18632/oncotarget.2971

Luo, M., Dennis, E. S., Berger, F., Peacock, W. J., & Chaudhury, A. (2005). MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. *Proceedings of the National Academy of Sciences*, *102*(48), 17531–17536. https://doi.org/10.1073/pnas.0508418102

Luscombe, N. M., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine*, *40*(4), 346–58. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11552348

Ma, Q., Geng, Y., Xu, W., Wu, Y., He, F., Shu, W., ... Li, M. (2010). The role of translationally controlled tumor protein in tumor growth and metastasis of colon adenocarcinoma cells. *Journal of Proteome Research*, *9*(1), 40–49. https://doi.org/10.1073/pnas.2335950100

MacDonald, J. I., & Dick, F. A. (2012). Posttranslational modifications of the retinoblastoma tumor suppressor protein as determinants of function. *Genes and Cancer*, *3*(11–12), 619–633. https://doi.org/10.1177/1947601912473305

MacDonald, S. M., Lichtenstein, L. M., Proud, D., Plaut, M., Naclerio, R. M., MacGlashan, D. W., & Kagey-Sobotka, A. (1987). Studies of IgE-dependent histamine releasing factors: Heterogeneity of IgE. *Journal of Immunology*, *139*(2), 506–512. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0023185832&partnerID=tZOtx3y1

MacDonald, S. M., Paznekas, W. A., & Jabs, E. W. (1999). Chromosomal localization of tumor protein, translationally-controlled 1 (TPT1) encoding the human histamine releasing factor (HRF) to 13q12-->q14. *Cytogenetics and Cell Genetics*, *84*(1–2), 128–9. https://doi.org/10.1159/000015238

MacDonald, S. M., Rafnar, T., Langdon, J., & Lichtenstein, L. M. (1995). Molecular identification of an IgE-dependent histamine-releasing factor. *Science (New York, N.Y.)*, *269*(5224), 688–90. https://doi.org/10.1126/science.7542803

Machado, H. L., Kittrell, F. S., Edwards, D., White, A. N., Atkinson, R. L., Rosen, J. M., ... Lewis, M. T. (2013). Separation by cell size enriches for mammary stem cell repopulation activity. *Stem Cells Translational Medicine*, *2*(3), 199–203. https://doi.org/10.5966/sctm.2012-0121

MacIver, N. J., Michalek, R. D., & Rathmell, J. C. (2013). Metabolic regulation of T

lymphocytes. *Annual Review of Immunology*, *31*, 259–83. https://doi.org/10.1146/annurev-immunol-032712-095956

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., & Ravikesavan, R. (2013). Gene duplication as a major force in evolution. *Journal of Genetics*, *92*(1), 155–61. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23640422

Mak, C. H., Poon, M. W., Lun, H. M., Kwok, P. Y., & Ko, R. C. (2007). Heat-inducible translationally controlled tumor protein of Trichinella pseudospiralis: cloning and regulation of gene expression. *Parasitology Research*, *100*(5), 1105–11. https://doi.org/10.1007/s00436-006-0373-y

Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2015). Archaeal Clusters of Orthologous Genes (arCOGs): An Update and Application for Analysis of Shared Features between Thermococcales, Methanococcales, and Methanobacteriales. *Life (Basel, Switzerland)*, *5*(1), 818–40. https://doi.org/10.3390/life5010818

Malumbres, M., & Barbacid, M. (2009). Cell cycle, CDKs and cancer: A changing paradigm. *Nature Reviews Cancer*, *9*(3), 153–166. https://doi.org/10.1038/nrc2602

Mamane, Y., Petroulakis, E., LeBacquer, O., & Sonenberg, N. (2006). mTOR, translation initiation and cancer. *Oncogene*, *25*(48), 6416–22. https://doi.org/10.1038/sj.onc.1209888

Mann, G. J., Musgrove, E. A., Fox, R. M., & Thelander, L. (1988). Ribonucleotide reductase M1 subunit in cellular proliferation, quiescence, and differentiation. *Cancer Research*, *48*(18), 5151–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3044582

Marchler-Bauer, A., Anderson, J. B., Derbyshire, M. K., DeWeese-Scott, C., Gonzales, N. R., Gwadz, M., ... Bryant, S. H. (2007). CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Research*, *35*(Database), D237–D240. https://doi.org/10.1093/nar/gkl951

Marshall, W. F. (2016). Cell Geometry: How Cells Count and Measure Size. *Annual Review of Biophysics*, *45*(1), 49–64. https://doi.org/10.1146/annurev-biophys-062215-010905

Marshall, W. F., Young, K. D., Swaffer, M., Wood, E., Nurse, P., Kimura, A., ... Roeder, A. H. K. (2012). What determines cell size? *BMC Biology*, *10*(101), 1–22. https://doi.org/10.1186/1741-7007-10-101

Martín-Benito, J., Boskovic, J., Gómez-Puertas, P., Carrascosa, J. L., Simons, C. T., Lewis, S. A., ... Valpuesta, J. M. (2002). Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT. *EMBO Journal*, *21*(23), 6377–6386. https://doi.org/10.1093/emboj/cdf640

Martín-Benito, J., Gómez-Reino, J., Stirling, P. C., Lundin, V. F., Gómez-Puertas, P., Boskovic, J., ... Valpuesta, J. M. (2007). Divergent Substrate-Binding Mechanisms Reveal an Evolutionary Specialization of Eukaryotic Prefoldin Compared to Its Archaeal Counterpart. *Structure*, *15*(1), 101–110. https://doi.org/10.1016/j.str.2006.11.006

Martin, D. E., Soulard, A., & Hall, M. N. (2004). TOR regulates ribosomal protein gene expression via PKA and the Forkhead transcription factor FHL1. *Cell*, *119*(7), 969–79. https://doi.org/10.1016/j.cell.2004.11.047

Martin, S. G., & Berthelot-Grosjean, M. (2009). Polar gradients of the DYRK-family kinase Pom1 couple cell length with the cell cycle. *Nature*, *459*(7248), 852–6.

https://doi.org/10.1038/nature08054

Martínez-Láinez, J. M., Moreno, D. F., Parisi, E., Clotet, J., & Aldea, M. (2018). Centromeric signaling proteins boost G1 cyclin degradation and modulate cell size in budding yeast. *PLoS Biology*, *16*(8), e2005388. https://doi.org/10.1371/journal.pbio.2005388

Martinez-Porchas, M., Villalpando-Canchola, E., Ortiz Suarez, L. E., & Vargas-Albores, F. (2017). How conserved are the conserved 16S-rRNA regions? *PeerJ*, *5*, e3036. https://doi.org/10.7717/peerj.3036

Masai, H., Taniyama, C., Ogino, K., Matsui, E., Kakusho, N., Matsumoto, S., ... Arai, K. I. (2006). Phosphorylation of MCM4 by Cdc7 kinase facilitates its interaction with Cdc45 on the chromatin. *Journal of Biological Chemistry*, *281*(51), 39249–39261. https://doi.org/10.1074/jbc.M608935200

Masura, S. S., Ahmad Parveez, G. K., & Eng Ti, L. L. (2011). Isolation and characterization of an oil palm constitutive promoter derived from a translationally control tumor protein (TCTP) gene. *Plant Physiology and Biochemistry*, *49*(7), 701–708. https://doi.org/10.1016/j.plaphy.2011.04.003

Matz, M. V., Frank, T. M., Marshall, N. J., Widder, E. A., & Johnsen, S. (2008). Giant deep-sea protist produces bilaterian-like traces. *Current Biology*, *18*(23), 1849–54. https://doi.org/10.1016/j.cub.2008.10.028

Mazhab-Jafari, M. T., Marshall, C. B., Ishiyama, N., Ho, J., Di Palma, V., Stambolic, V., & Ikura, M. (2012). An autoinhibited noncanonical mechanism of GTP hydrolysis by Rheb maintains mTORC1 homeostasis. *Structure*, *20*(9), 1528–1539. https://doi.org/10.1016/j.str.2012.06.013

McDowall, M. D., Harris, M. A., Lock, A., Rutherford, K., Staines, D. M., Bähler, J. Ü., ... Wood, V. (2015). PomBase 2015: Updates to the fission yeast database. *Nucleic Acids Research*, *43*(D1), D656–D661. https://doi.org/10.1093/nar/gku1040

McGary, K. L., Park, T. J., Woods, J. O., Cha, H. J., Wallingford, J. B., & Marcotte, E. M. (2010). Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proceedings of the National Academy of Sciences*, *107*(14), 6544–6549. https://doi.org/10.1073/pnas.0910200107

Meek, D. W. (2009). Tumour suppression by p53: a role for the DNA damage response? *Nature Reviews. Cancer*, *9*(10), 714–23. https://doi.org/10.1038/nrc2716

Meinke, D. W. (1998). Arabidopsis thaliana: A Model Plant for Genome Analysis. *Science*, *282*(5389), 662–682. https://doi.org/10.1126/science.282.5389.662

Meldal, B. H. M. M., Forner-Martinez, O., Costanzo, M. C., Dana, J., Demeter, J., Dumousseau, M., ... Orchard, S. (2015). The complex portal - An encyclopaedia of macromolecular complexes. *Nucleic Acids Research*, *43*(D1), D479–D484. https://doi.org/10.1093/nar/gku975

Meziane, E. K., Randle, S. J., Nelson, D. E., Lomonosov, M., & Laman, H. (2011). Knockdown of Fbxo7 reveals its regulatory role in proliferation and differentiation of haematopoietic precursor cells. *Journal of Cell Science*, *124*(Pt 13), 2175–86. https://doi.org/10.1242/jcs.080465

Miettinen, T. P., & Björklund, M. (2016). Cellular Allometry of Mitochondrial Functionality Establishes the Optimal Cell Size. *Developmental Cell*, *39*(3), 370–382. https://doi.org/10.1016/j.devcel.2016.09.004

Milne, D., Campbell, D., Caudwell, F., & Meek, D. (1994). Phosphorylation of the tumor suppressor protein p53 by mitogen-activated protein kinases. *Journal of Biological Chemistry*, *269*(12), 9253–9260. Retrieved from http://www.jbc.org/content/269/12/9253.short

Miluzio, A., Ricciardi, S., Manfrini, N., Alfieri, R., Oliveto, S., Brina, D., & Biffo, S. (2016). Translational control by mTOR-independent routes: how eIF6 organizes metabolism. *Biochemical Society Transactions*, *44*(6), 1667–1673. https://doi.org/10.1042/BST20160179

Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., ... Finn, R. D. (2015). The InterPro protein families database: The classification resource after 15 years. *Nucleic Acids Research*, *43*(D1), D213–D221. https://doi.org/10.1093/nar/gku1243

Mittal, S., Mallikarjuna, M. G., Rao, A. R., Jain, P. A., Dash, P. K., & Thirunavukkarasu, N. (2017). Comparative Analysis of CDPK Family in Maize, Arabidopsis, Rice, and Sorghum Revealed Potential Targets for Drought Tolerance Improvement. *Frontiers in Chemistry*, *5*(December), 1–17. https://doi.org/10.3389/fchem.2017.00115

Miyazawa, M., Tashiro, E., Kitaura, H., Maita, H., Suto, H., Iguchi-Ariga, S. M. M., & Ariga, H. (2011). Prefoldin subunits are protected from ubiquitin-proteasome system-mediated degradation by forming complex with other constituent subunits. *Journal of Biological Chemistry*, *286*(22), 19191–19203. https://doi.org/10.1074/jbc.M110.216259

Miyoshi, A., Kito, K., Aramoto, T., Abe, Y., Kobayashi, N., & Ueda, N. (2003). Identification of CGI-121, a novel PRPK (p53-related protein kinase)-binding protein. *Biochemical and Biophysical Research Communications*, *303*(2), 399–405. https://doi.org/10.1016/S0006-291X(03)00333-4

Mizukami, Y. (2001). A matter of size: developmental control of organ size in plants. *Current Opinion in Plant Biology*, *4*(6), 533–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11641070

Mizukami, Y., & Fischer, R. L. (2000). Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. *Proceedings of the National Academy of Sciences*, *97*(2), 942–947. https://doi.org/10.1073/pnas.97.2.942

Montagnoli, A., Bosotti, R., Villa, F., Rialland, M., Brotherton, D., Mercurio, C., ... Santocanale, C. (2002). Drf1, a novel regulatory subunit for human Cdc7 kinase. *EMBO Journal*, *21*(12), 3171–3181. https://doi.org/10.1093/emboj/cdf290

Moore, J. K., Scheinman, R. I., & Bellgrau, D. (2001). The identification of a novel T cell activation state controlled by a diabetogenic gene. *Journal of Immunology (Baltimore, Md. : 1950)*, *166*(1), 241–8. https://doi.org/10.4049/jimmunol.166.1.241

Moreno, S., Hayles, J., & Nurse, P. (1989). Regulation of p34cdc2 protein kinase during mitosis. *Cell*, *58*(2), 361–72. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2665944

Moretto, F., Sagot, I., Daignan-Fornier, B., & Pinson, B. (2013). A pharmaco-epistasis strategy reveals a new cell size controlling pathway in yeast. *Molecular Systems Biology*, *9*(707), 707. https://doi.org/10.1038/msb.2013.60

Mori, K., Maeda, Y., Kitaura, H., Taira, T., Iguchi-Ariga, S. M. M., & Ariga, H.

(1998). MM-1, a novel c-Myc-associating protein that represses transcriptional activity of c-Myc. *Journal of Biological Chemistry*, *273*(45), 29794–29800. https://doi.org/10.1074/jbc.273.45.29794

Moseley, J. B., Mayeux, A., Paoletti, A., & Nurse, P. (2009). A spatial gradient coordinates cell size and mitotic entry in fission yeast. *Nature*, *459*(7248), 857–60. https://doi.org/10.1038/nature08074

Mousnier, A., Kubat, N., Massias-Simon, A., Segeral, E., Rain, J.-C., Benarous, R., ... Dargemont, C. (2007). von Hippel Lindau binding protein 1-mediated degradation of integrase affects HIV-1 gene expression at a postintegration step. *Proceedings of the National Academy of Sciences*, *104*(34), 13615–13620. https://doi.org/10.1073/pnas.0705162104

Muller, J., Szklarczyk, D., Julien, P., Letunic, I., Roth, A., Kuhn, M., ... Bork, P. (2010). eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Research*, *38*(Database issue), D190-5. https://doi.org/10.1093/nar/gkp951

Mulligan, G. J., Wong, J., & Jacks, T. (1998). p130 is dispensable in peripheral T lymphocytes: evidence for functional compensation by p107 and pRB. *Molecular and Cellular Biology*, *18*(1), 206–20. https://doi.org/10.1128/MCB.18.1.206

Murzin, A. G., Brenner, S. E., Hubbard, T., & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology*, *247*(4), 536–40. https://doi.org/10.1006/jmbi.1995.0159

Nakamura, H., Morita, T., Masaki, S., & Yoshida, S. (1984). Intracellular localization and metabolism of DNA polymerase alpha in human cells visualized with monoclonal antibody. *Experimental Cell Research*, *151*(1), 123–33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6421608

Nakamura, S., Roth, J. A., & Mukhopadhyay, T. (2000). Multiple lysine mutations in the C-terminal domain of p53 interfere with MDM2-dependent protein degradation and ubiquitination. *Molecular and Cellular Biology*, *20*(24), 9391–9398. https://doi.org/10.1128/MCB.20.24.9391-9398.2000

Nakashima, A., Sato, T., & Tamanoi, F. (2010). Fission yeast TORC1 regulates phosphorylation of ribosomal S6 proteins in response to nutrients and its activity is inhibited by rapamycin. *Journal of Cell Science*, *123*(5), 777–786. https://doi.org/10.1242/jcs.060319

Nakayama, K., Ishida, N., Shirane, M., Inomata, A., Inoue, T., Shishido, N., ... Beach, D. (1996). Mice lacking p27(Kip1) display increased body size, multiple organ hyperplasia, retinal dysplasia, and pituitary tumors. *Cell*, *85*(5), 707–20. https://doi.org/10.1016/s0092-8674(00)81237-4

Narita, N. N., Moore, S., Horiguchi, G., Kubo, M., Demura, T., Fukuda, H., ... Tsukaya, H. (2004). Overexpression of a novel small peptide ROTUNDIFOLIA4 decreases cell proliferation and alters leaf shape in Arabidopsis thaliana. *The Plant Journal : For Cell and Molecular Biology*, *38*(4), 699–713. https://doi.org/10.1111/j.1365-313X.2004.02078.x

Nemoto, K., Seto, T., Takahashi, H., Nozawa, A., Seki, M., Shinozaki, K., ... Sawasaki, T. (2011). Autophosphorylation profiling of Arabidopsis protein kinases using the cell-free system. *Phytochemistry*, *72*(10), 1136–1144.

https://doi.org/10.1016/j.phytochem.2011.02.029
Neufeld, T. P., de la Cruz, a F., Johnston, L. a, & Edgar, B. A. (1998). Coordination of growth and cell division in the Drosophila wing. *Cell*, *93*(7), 1183–93. https://doi.org/10.1016/S0092-8674(00)81462-2
Neufeld, T. P., & Edgart, B. A. (1998). Connections between growth and the cell cycle. *Current Opinion in Cell Biology*, *10*(10), 784–790.
Neumann, B., Walter, T., Hériché, J.-K., Bulkescher, J., Erfle, H., Conrad, C., ... Ellenberg, J. (2010). Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, *464*(7289), 721–7. https://doi.org/10.1038/nature08869
Neumann, F. R., & Nurse, P. (2007). Nuclear size control in fission yeast. *The Journal of Cell Biology*, *179*(4), 593–600. https://doi.org/10.1083/jcb.200708054
Ngoenkam, J., Schamel, W. W., & Pongcharoen, S. (2017). Selected signalling proteins recruited to the T-cell receptor-CD3 complex. *Immunology*, 1–9. https://doi.org/10.1111/imm.12809
Nieduszynski, C. A., Blow, J. J., & Donaldson, A. D. (2005). The requirement of yeast replication origins for pre-replication complex proteins is modulated by transcription. *Nucleic Acids Research*, *33*(8), 2410–2420. https://doi.org/10.1093/nar/gki539
Nolo, R., Morrison, C. M., Tao, C., Zhang, X., & Halder, G. (2006). The bantam MicroRNA Is a Target of the Hippo Tumor-Suppressor Pathway. *Current Biology*, *16*(19), 1895–1904. https://doi.org/10.1016/j.cub.2006.08.057
Nurse, P., & Thuriaux, P. (1980). Regulatory genes controlling mitosis in the fission yeast Schizosaccharomyces pombe. *Genetics*, *96*(3), 627–37. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7262540
Nurse, P., Thuriaux, P., & Nasmyth, K. (1976). Genetic control of the cell division cycle in the fission yeast Schizosaccharomyces pombe. *MGG Molecular & General Genetics*, *146*(2), 167–178. https://doi.org/10.1007/BF00268085
O'Brien, K. P., Remm, M., & Sonnhammer, E. L. L. (2005). Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Research*, *33*(Database issue), D476-80. https://doi.org/10.1093/nar/gki107
O'Malley, R. C., Barragan, C. C., & Ecker, J. R. (2015). A user's guide to the arabidopsis T-DNA insertion mutant collections. *Plant Functional Genomics: Methods and Protocols: Second Edition*, 323–342. https://doi.org/10.1007/978-1-4939-2444-8_16
O'Malley, R. C., & Ecker, J. R. (2010). Linking genotype to phenotype using the Arabidopsis unimutant collection. *Plant Journal*, *61*(6), 928–940. https://doi.org/10.1111/j.1365-313X.2010.04119.x
Oehlmann, M., Score, A. J., & Blow, J. J. (2004). The role of Cdc6 in ensuring complete genome licensing and S phase checkpoint activation. *Journal of Cell Biology*, *165*(2), 181–190. https://doi.org/10.1083/jcb.200311044
Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *27*(1), 29–34. https://doi.org/10.1093/nar/27.1.29
Onnis, A., Finetti, F., & Baldari, C. T. (2016). Vesicular trafficking to the immune synapse: How to assemble receptor-tailored pathways from a basic building set. *Frontiers in Immunology*, *7*(FEB), 1–9.

https://doi.org/10.3389/fimmu.2016.00050

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., ... Hermjakob, H. (2014). The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research*, *42*(Database issue), D358-63. https://doi.org/10.1093/nar/gkt1115

Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., & Thornton, J. M. (1997). CATH--a hierarchic classification of protein domain structures. *Structure (London, England : 1993)*, *5*(8), 1093–108. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9309224

Orr, S. J., Boutz, D. R., Wang, R., Chronis, C., Lea, N. C., Thayaparan, T., ... Thomas, N. S. B. (2012). Proteomic and protein interaction network analysis of human T lymphocytes during cell-cycle entry. *Molecular Systems Biology*, *8*(573), 573. https://doi.org/10.1038/msb.2012.5

Orr, S. J., Gaymes, T., Ladon, D., Chronis, C., Czepulkowski, B., Wang, R., ... Thomas, N. S. B. (2010). Reducing MCM levels in human primary T cells during the G0→G1 transition causes genomic instability during the first cell cycle. *Oncogene*, *29*(26), 3803–3814. https://doi.org/10.1038/onc.2010.138

Ostlund, G., Schmitt, T., Forslund, K., Köstler, T., Messina, D. N., Roopra, S., ... Sonnhammer, E. L. L. (2010). InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Research*, *38*(Database issue), D196-203. https://doi.org/10.1093/nar/gkp931

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., ... Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, *42*(D1), 206–214. https://doi.org/10.1093/nar/gkt1226

Oyama, T., Shimura, Y., & Okada, K. (1997). The Arabidopsis HY5 gene encodes a bZIP protein that regulates stimulus- induced development of root and hypocotyl. *Genes and Development*, *11*(22), 2983–2995. https://doi.org/10.1101/gad.11.22.2983

Pagani, F., Raponi, M., & Baralle, F. E. (2005). Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proceedings of the National Academy of Sciences*, *102*(18), 6368–6372. https://doi.org/10.1073/pnas.0502288102

Pan, D. (2010). The hippo signaling pathway in development and cancer. *Developmental Cell*, *19*(4), 491–505. https://doi.org/10.1016/j.devcel.2010.09.011

Pardee, A. B. (1974). A restriction point for control of normal animal cell proliferation. *Proceedings of the National Academy of Sciences of the United States of America*, *71*(4), 1286–90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4524638

Pavletich, N. P. (1999). Mechanisms of cyclin-dependent kinase regulation: Structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors. *Journal of Molecular Biology*, *287*(5), 821–828. https://doi.org/10.1006/jmbi.1999.2640

Pearson, W. R. (2013). An Introduction to Sequence Similarity ("Homology") Searching. *Current Protocols in Bioinformatics*, *42*(1), 3.1.1-3.1.8. https://doi.org/10.1002/0471250953.bi0301s42

Pearson, W. R., & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of*

America, *85*(8), 2444–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3162770

Pedruzzi, I., Dubouloz, F., Cameroni, E., Wanke, V., Roosen, J., Winderickx, J., & De Virgilio, C. (2003). TOR and PKA Signaling Pathways Converge on the Protein Kinase Rim15 to Control Entry into G0. *Molecular Cell*, *12*(6), 1607–1613. https://doi.org/10.1016/S1097-2765(03)00485-4

Penicud, K., & Behrens, A. (2014). DMAP1 is an essential regulator of ATM activity and function. *Oncogene*, *33*(4), 525–531. https://doi.org/10.1038/onc.2012.597

Perea-Resa, C., Rodríguez-Milla, M. A., Iniesto, E., Rubio, V., & Salinas, J. (2017). Prefoldins Negatively Regulate Cold Acclimation in Arabidopsis thaliana by Promoting Nuclear Proteasome-Mediated HY5 Degradation. *Molecular Plant*, *10*(6), 791–804. https://doi.org/10.1016/j.molp.2017.03.012

Perillo, N. L., Walford, R. L., Newman, M. A., & Effros, R. B. (1989). Human T lymphocytes possess a limited in vitro life span. *Experimental Gerontology*, *24*(3), 177–87. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2786475

Peterson, D., Lee, J., Lei, X. C., Forrest, W. F., Davis, D. P., Jackson, P. K., & Belmont, L. D. (2010). A chemosensitization screen identifies TP53RK, a kinase that restrains apoptosis after mitotic stress. *Cancer Research*, *70*(15), 6325–6335. https://doi.org/10.1158/0008-5472.CAN-10-0015

Pevsner, J. (2015). *Bioinformatics and Functional Genomics*. *Briefings in Functional Genomics and Proteomics* (Vol. 3). Singapore: Wiley. Retrieved from https://www.wiley.com/en-us/Bioinformatics+and+Functional+Genomics%2C+3rd+Edition-p-9781118581780

Pollizzi, K. N., Waickman, A. T., Patel, C. H., Sun, I. H., & Powell, J. D. (2015). Cellular size as a means of tracking mTOR activity and cell fate of CD4+ T cells upon antigen recognition. *PLoS ONE*, *10*(4), 1–22. https://doi.org/10.1371/journal.pone.0121710

Porciello, N., & Tuosto, L. (2016). CD28 costimulatory signals in T lymphocyte activation: Emerging functions beyond a qualitative and quantitative support to TCR signalling. *Cytokine and Growth Factor Reviews*, *28*, 11–19. https://doi.org/10.1016/j.cytogfr.2016.02.004

Porter, L. A., & Donoghue, D. J. (2003). Cyclin B1 and CDK1: nuclear localization and upstream regulators. *Progress in Cell Cycle Research*, *5*, 335–47. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14593728

Poux, S., & Gaudet, P. (2017). Best Practices in Manual Annotation with the Gene Ontology (pp. 41–54). https://doi.org/10.1007/978-1-4939-3743-1_4

Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., ... Bork, P. (2014). eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Research*, *42*(Database issue), D231-9. https://doi.org/10.1093/nar/gkt1253

Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., ... Bork, P. (2012). eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Research*, *40*(Database issue), D284-9. https://doi.org/10.1093/nar/gkr1060

Preston, R. A., Manolson, M. F., Becherer, K., Weidenhammer, E., Kirkpatrick, D., Wright, R., & Jones, E. W. (1991). Isolation and characterization of PEP3, a gene

required for vacuolar biogenesis in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *11*(12), 5801–5812. https://doi.org/10.1128/MCB.11.12.5801.Updated

Provan, D., Singer, C. R. J., Baglin, T., & Lilleyman, J. (2004). *Oxford Handbook of Clinical Haematology* (second ed.). Oxford: Oxford University Press. Retrieved from http://lib.medilam.ac.ir/Portals/81/ebook/hemato/Oxford Handbook Of Clinical Haematology.pdf?ver=1395-04-24-202408-320

Provart, N. J., Alonso, J., Assmann, S. M., Bergmann, D., Brady, S. M., Brkljacic, J., ... McCourt, P. (2016). 50 years of Arabidopsis research: highlights and future directions. *The New Phytologist*, *209*(3), 921–44. https://doi.org/10.1111/nph.13687

Rathmell, J. C., Heiden, M. G. V., Harris, M. H., Frauwirth, K. A., & Thompson, C. B. (2000). In the absence of extrinsic signals, nutrient utilization by lymphocytes is insufficient to maintain either cell size or viability. *Molecular Cell*, *6*(3), 683–692. https://doi.org/10.1016/S1097-2765(00)00066-6

Reichert, P., Reinhardt, R. L., Ingulli, E., & Jenkins, M. K. (2001). Cutting edge: in vivo identification of TCR redistribution and polarized IL-2 production by naive CD4 T cells. *Journal of Immunology (Baltimore, Md. : 1950)*, *166*(7), 4278–81. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11254679

Ren, S., & Rollins, B. J. (2004). Cyclin C/Cdk3 promotes Rb-dependent G0 exit. *Cell*, *117*(2), 239–251. https://doi.org/10.1016/S0092-8674(04)00300-9

Rhee, S. Y., & Mutwil, M. (2014). Towards revealing the functions of all genes in plants. *Trends in Plant Science*. https://doi.org/10.1016/j.tplants.2013.10.006

Rho, S. B., Lee, J. H., Park, M. S., Byun, H. J., Kang, S., Seo, S. S., ... Park, S. Y. (2011). Anti-apoptotic protein TCTP controls the stability of the tumor suppressor p53. *FEBS Letters*, *585*(1), 29–35. https://doi.org/10.1016/j.febslet.2010.11.014

Riera, A., Barbon, M., Noguchi, Y., Reuter, L. M., Schneider, S., & Speck, C. (2017). From structure to mechanism — understanding initiation of DNA replication. *Genes & Development*, *31*(11), 1073–1088. https://doi.org/10.1101/gad.298232.117.

Rigden, D. J., & Fernández, X. M. (2018). The 2018 Nucleic Acids Research database issue and the online molecular biology database collection. *Nucleic Acids Research*, *46*(D1), D1–D7. https://doi.org/10.1093/nar/gkx1235

Rinnerthaler, M., Jarolim, S., Heeren, G., Palle, E., Perju, S., Klinger, H., ... Laun, P. (2006). MMI1 (YKL056c, TMA19), the yeast orthologue of the translationally controlled tumor protein (TCTP) has apoptotic functions and interacts with both microtubules and mitochondria. *Biochimica et Biophysica Acta*, *1757*(5–6), 631–8. https://doi.org/10.1016/j.bbabio.2006.05.022

Risal, S., Adhikari, D., & Liu, K. (2016). Animal Models for Studying the In Vivo Functions of Cell Cycle CDKs. *Methods in Molecular Biology (Clifton, N.J.)*, *1336*, 155–66. https://doi.org/10.1007/978-1-4939-2926-9_13

Rittling, S. R., Brooks, K. M., Cristofalo, V. J., & Baserga, R. (1986). Expression of cell cycle-dependent genes in young and senescent WI-38 fibroblasts. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(10), 3316–20. https://doi.org/10.1073/pnas.83.10.3316

Rodríguez-Milla, M. A., & Salinas, J. (2009). Prefoldins 3 and 5 play an essential role in Arabidopsis tolerance to salt stress. *Molecular Plant*, *2*(3), 526–534.

https://doi.org/10.1093/mp/ssp016

Rodríguez-Tarduchy, G., & López-Rivas, A. (1989). Phorbol esters inhibit apoptosis in IL-2-dependent T lymphocytes. *Biochemical and Biophysical Research Communications*, *164*(3), 1069–75. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2590187

Rojo, E., Zouhar, J., Kovaleva, V., Hong, S., & Raikhel, N. V. (2003). The AtC-VPS protein complex is localized to the tonoplast and the prevacuolar compartment in arabidopsis. *Molecular Biology of the Cell*, *14*(2), 361–9. https://doi.org/10.1091/mbc.E02-08-0509

Rothenberg, E. V. (2014). The chromatin landscape and transcription factors in T cell programming. *Trends in Immunology*, *35*(5), 195–204. https://doi.org/10.1016/j.it.2014.03.001

Rupes, I. (2002). Checking cell size in yeast. *Trends in Genetics : TIG*, *18*(9), 479–85. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12175809

Russell, P., & Nurse, P. (1987). Negative regulation of mitosis by wee1+, a gene encoding a protein kinase homolog. *Cell*, *49*(4), 559–67. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3032459

Sahlan, M., Zako, T., & Yohda, M. (2018). Prefoldin, a jellyfish-like molecular chaperone: functional cooperation with a group II chaperonin and beyond. *Biophysical Reviews*. https://doi.org/10.1007/s12551-018-0400-0

Sakaguchi, K., Herrera, J. E., Saito, S., Miki, T., Bustin, M., Vassilev, A., ... Appella, E. (1998). DNA damage activates p53 through a phosphorylation-acetylation cascade. *Genes and Development*, *12*(18), 2831–2841. https://doi.org/10.1101/gad.12.18.2831

Sakaguchi, K., Saito, S., Higashimoto, Y., Roy, S., Anderson, C. W., & Appella, E. (2000). Damage-mediated phosphorylation of human p53 threonine 18 through a cascade mediated by a casein 1-like kinase. Effect on MDM2 binding. *Journal of Biological Chemistry*, *275*(13), 9278–9283. https://doi.org/10.1074/jbc.275.13.9278

Salem, M. A., Li, Y., Wiszniewski, A., & Giavalisco, P. (2017). Regulatory-associated protein of TOR (RAPTOR) alters the hormonal and metabolic composition of Arabidopsis seeds, controlling seed morphology, viability and germination potential. *Plant Journal*, *92*(4), 525–545. https://doi.org/10.1111/tpj.13667

Sanchez, J. C., Schaller, D., Ravier, F., Golaz, O., Jaccoud, S., Belet, M., ... Hochstrasser, D. (1997). Translationally controlled tumor protein: a protein identified in several nontumoral cells including erythrocytes. *Electrophoresis*, *18*(1), 150–5. https://doi.org/10.1002/elps.1150180127

Sansores-Garcia, L., Bossuyt, W., Wada, K. I., Yonemura, S., Tao, C., Sasaki, H., & Halder, G. (2011). Modulating F-actin organization induces organ growth by affecting the Hippo pathway. *EMBO Journal*, *30*(12), 2325–2335. https://doi.org/10.1038/emboj.2011.157

Scherer, L. J., & Rossi, J. J. (2003). Approaches for the sequence-specific knockdown of mRNA. *Nature Biotechnology*, *21*(12), 1457–1465. https://doi.org/10.1038/nbt915

Schlueter, S. D., Dong, Q., & Brendel, V. (2003). GeneSeqer@PlantGDB: Gene structure prediction in plant genomes. *Nucleic Acids Research*, *31*(13), 3597–3600. https://doi.org/10.1093/nar/gkg533

Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., ... Lohmann, J. U. (2005). A gene expression map of Arabidopsis thaliana development. *Nature Genetics*, *37*(5), 501–6. https://doi.org/10.1038/ng1543

Schopf, J. W., Kudryavtsev, A. B., Agresti, D. G., Wdowiak, T. J., & Czaja, A. D. (2002). Laser–Raman imagery of Earth's earliest fossils. *Nature*, *416*(6876), 73–76. https://doi.org/10.1038/416073a

Schopfer, P. (2006). Biomechanics of plant growth. *American Journal of Botany*, *93*(10), 1415–25. https://doi.org/10.3732/ajb.93.10.1415

Schruff, M. C. (2005). The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. *Development*, *133*(2), 251–261. https://doi.org/10.1242/dev.02194

Schulz, E., Tohge, T., Zuther, E., Fernie, A. R., & Hincha, D. K. (2015). Natural variation in flavonol and anthocyanin metabolism during cold acclimation in Arabidopsis thaliana accessions. *Plant, Cell and Environment*, *38*(8), 1658–1672. https://doi.org/10.1111/pce.12518

Schumann, K., Lin, S., Boyer, E., Simeonov, D. R., Subramaniam, M., Gate, R. E., ... Marson, A. (2015). Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(33), 10437–42. https://doi.org/10.1073/pnas.1512503112

Schwartz, R. H. (2003). T Cell Anergy. *Annual Review of Immunology*, *21*(1), 305–334. https://doi.org/10.1146/annurev.immunol.21.120601.141110

Schwarz, B. A., & Bhandoola, A. (2006). Trafficking from the bone marrow to the thymus: A prerequisite for thymopoiesis. *Immunological Reviews*, *209*, 47–57. https://doi.org/10.1111/j.0105-2896.2006.00350.x

Shahinian, A., Pfeffer, K., Lee, K. P., Kündig, T. M., Kishihara, K., Wakeham, A., ... Mak, T. W. (1993). Differential T cell costimulatory requirements in CD28-deficient mice. *Science (New York, N.Y.)*, *261*(5121), 609–12. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7688139

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–504. https://doi.org/10.1101/gr.1239303

Sherr, C. J. (1994). G1 phase progression: Cycling on cue. *Cell*, *79*(4), 551–555. https://doi.org/10.1016/0092-8674(94)90540-1

Sherr, C. J., & Roberts, J. M. (1995). Inhibitors of mammalian cyclin-dependent kinases. *Genes & Development*, 1149–1163. https://doi.org/10.1101/gad.9.10.1149

Sherr, C. J., & Roberts, J. M. (1999). CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes & Development*, *13*(12), 1501–12. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10385618

Shi, Y. F., Szalay, M. G., Paskar, L., Sahai, B. M., Boyer, M., Singh, B., & Green, D. R. (1990). Activation-induced cell death in T cell hybridomas is due to apoptosis. Morphologic aspects and DNA fragmentation. *Journal of Immunology (Baltimore, Md. : 1950)*, *144*(9), 3326–33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1691753

Shieh, S. Y., Ahn, J., Tamai, K., Taya, Y., & Prives, C. (2000). The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate, p53 at multiple DNA

damage-inducible sites. *Genes and Development*, *14*(3), 289–300. https://doi.org/10.1101/gad.14.3.289

Shieh, S. Y., Ikeda, M., Taya, Y., & Prives, C. (1997). DNA damage-induced phosphorylation of p53 alleviates inhibition by MDM2. *Cell*, *91*(3), 325–334. https://doi.org/10.1016/S0092-8674(00)80416-X

Siegers, K., Waldmann, T., Leroux, M. R., Grein, K., Shevchenko, A., Schiebel, E., & Hartl, F. U. (1999). Compartmentation of protein folding in vivo: Sequestration of non-native polypeptide by the chaperonin-GimC system. *EMBO Journal*, *18*(1), 75–84. https://doi.org/10.1093/emboj/18.1.75

Siegert, R., Leroux, M. R., Scheufler, C., Hartl, F. U., & Moarefi, I. (2000). Structure of the Molecular Chaperone Prefoldin. *Cell*, *103*(4), 621–632. https://doi.org/10.1016/S0092-8674(00)00165-3

Simanis, V., & Nurse, P. (1986). The cell cycle control gene cdc2+ of fission yeast encodes a protein kinase potentially regulated by phosphorylation. *Cell*, *45*(2), 261–268. https://doi.org/10.1016/0092-8674(86)90390-9

Simoni, Y., Diana, J., Ghazarian, L., Beaudoin, L., & Lehuen, A. (2013). Therapeutic manipulation of natural killer (NK)T cells in autoimmunity: Are we close to reality? *Clinical and Experimental Immunology*, *171*(1), 8–19. https://doi.org/10.1111/j.1365-2249.2012.04625.x

Sipiczki, M. (1995). Phylogenesis of fission yeasts. Contradictions surrounding the origin of a century old genus. *Antonie van Leeuwenhoek*, *68*(2), 119–149. https://doi.org/10.1007/BF00873099

Sipiczki, M. (2000). Where does fission yeast sit on the tree of life? *Genome Biology*, *1*(2), REVIEWS1011. https://doi.org/10.1186/gb-2000-1-2-reviews1011

Sippel, K. H., Venkatakrishnan, B., Boehlein, S. K., Quirit, J. G., Govindasamy, L., Agbandje-mckenna, M., ... Mckenna, R. (2012). Insights into Mycoplasma genitalium Metabolism Revealed by the Structure of MG289, an Extracytoplasmic Thiamine Binding Lipoprotein. *Proteins*, *79*(2), 528–536. https://doi.org/10.1002/prot.22900.Insights

Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular subsequences. *Molecular Biology*, *147*, 195–197. https://doi.org/10.1016/0022-2836(81)90087-5

Springer, P. S., Holding, D. R., Groover, A., Yordan, C., & Martienssen, R. A. (2000). The essential Mcm7 protein PROLIFERA is localized to the nucleus of dividing cells during the G(1) phase and is required maternally for early Arabidopsis development. *Development*, *127*(9), 1815–1822. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10751170

Srinivasan, M., Mehta, P., Yu, Y., Prugar, E., Koonin, E. V., Karzai, A. W., & Sternglanz, R. (2011). The highly conserved KEOPS/EKC complex is essential for a universal tRNA modification, t6A. *EMBO Journal*, *30*(5), 873–881. https://doi.org/10.1038/emboj.2010.343

St Pierre, S. E., Ponting, L., Stefancsik, R., & McQuilton, P. (2014). FlyBase 102--advanced approaches to interrogating FlyBase. *Nucleic Acids Research*, *42*(Database issue), D780-8. https://doi.org/10.1093/nar/gkt1092

Standerholen, F. B., Myromslien, F. D., Kommisrud, E., Ropstad, E., & Waterhouse, K. E. (2014). Comparison of electronic volume and forward scatter principles of

cell selection using flow cytometry for the evaluation of acrosome and plasma membrane integrity of bull spermatozoa. *Cytometry. Part A : The Journal of the International Society for Analytical Cytology*, *85*(8), 719–28. https://doi.org/10.1002/cyto.a.22474

Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., & Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Research*, *34*(Database issue), D535-9. https://doi.org/10.1093/nar/gkj109

Stocchetto, S., Marin, O., Carignani, G., & Pinna, L. A. (1997). Biochemical evidence that Saccharomyces cerevisiae YGR262c gene, required for normal growth, encodes a novel Ser/Thr-specific protein kinase. *FEBS Letters*, *414*(1), 171–175. https://doi.org/10.1016/S0014-5793(97)00980-0

Stoeber, K., Tlsty, T. D., Happerfield, L., Thomas, G. a, Romanov, S., Bobrow, L., ... Williams, G. H. (2001). DNA replication licensing and human cell proliferation. *Journal of Cell Science*, *114*(Pt 11), 2027–2041. https://doi.org/10.1073/pnas.93.24.13742

Stone, S. L., Troy, A., Herschleb, J., Kraft, E., Callis, J., & Hauksdo, H. (2005). Functional Analysis of the RING-Type Ubiquitin Ligase Family of Arabidopsis 1 [w]. *Plant Physiology*, *137*(January), 13–30. https://doi.org/10.1104/pp.104.052423.carrying

Su, S. H., & Krysan, P. J. (2016). A double-mutant collection targeting MAP kinase related genes in Arabidopsis for studying genetic interactions. *Plant Journal*, *88*(5), 867–878. https://doi.org/10.1111/tpj.13292

Su, T. T., & O'Farrell, P. H. (1998). Size control: cell proliferation does not equal growth. *Current Biology : CB*, *8*(19), R687-9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9768354

Sun, B., & Zhang, Y. (2014). *T Helper Cell Differentiation and Their Function*. (B. Sun, Ed.) (Vol. 841). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-9487-9

Sundaresan, V. (2005). Control of seed size in plants. *Proc Natl Acad Sci U S A*, *102*(50), 17887–17888. https://doi.org/10.1073/pnas.0509021102

Surh, C. D., & Sprent, J. (2008). Homeostasis of Naive and Memory T Cells. *Immunity*, *29*(6), 848–862. https://doi.org/10.1016/j.immuni.2008.11.002

Susini, L., Besse, S., Duflaut, D., Lespagnol, A., Beekman, C., Fiucci, G., ... Telerman, A. (2008). TCTP protects from apoptotic cell death by antagonizing bax function. *Cell Death and Differentiation*, *15*(8), 1211–1220. https://doi.org/10.1038/cdd.2008.18

Sveiczer, A., Novak, B., & Mitchison, J. M. (1996). The size control of fission yeast revisited. *Journal of Cell Science*, *109 (Pt 1*, 2947–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9013342

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T. Z., Garcia-Hernandez, M., Foerster, H., ... Huala, E. (2008). The Arabidopsis Information Resource (TAIR): Gene structure and function annotation. *Nucleic Acids Research*, *36*(SUPPL. 1). https://doi.org/10.1093/nar/gkm965

Taheri-Araghi, S., Bradde, S., Sauls, J. T., Hill, N. S., Levin, P. A., Paulsson, J., ... Jun, S. (2014). Cell-Size Control and Homeostasis in Bacteria. *Current Biology*, *25*(3), 385–391. https://doi.org/10.1016/j.cub.2014.12.009

Takada, K., & Jameson, S. C. (2009). Naive T cell homeostasis: From awareness of

space to a sense of place. *Nature Reviews Immunology*, *9*(12), 823–832. https://doi.org/10.1038/nri2657

Takemoto, K., Ebine, K., Askani, J. C., Krüger, F., Gonzalez, Z. A., Ito, E., ... Ueda, T. (2018). Distinct sets of tethering complexes, SNARE complexes, and Rab GTPases mediate membrane fusion at the vacuole in Arabidopsis. *Proceedings of the National Academy of Sciences*, *115*(10), 201717839. https://doi.org/10.1073/pnas.1717839115

Tamura, K., & Hara-Nishimura, I. (2014). Functional insights of nucleocytoplasmic transport in plants. *Frontiers in Plant Science*, *5*(April), 1–10. https://doi.org/10.3389/fpls.2014.00118

Tani, T., Shimada, H., Kato, Y., & Tsunoda, Y. (2007). Bovine Oocytes with the Potential to Reprogram Somatic Cell Nuclei Have a Unique 23-kDa Protein, Phosphorylated Transcriptionally Controlled Tumor Protein (TCTP). *Cloning and Stem Cells*, *9*(2), 267–280. https://doi.org/10.1089/clo.2006.0072

Tao, J.-J., Cao, Y.-R., Chen, H.-W., Wei, W., Li, Q.-T., Ma, B., ... Zhang, J.-S. (2015). Tobacco Translationally Controlled Tumor Protein Interacts with Ethylene Receptor Tobacco Histidine Kinase1 and Enhances Plant Growth through Promotion of Cell Proliferation. *Plant Physiology*, *169*(1), 96–114. https://doi.org/10.1104/pp.15.00355

Tapon, N., Harvey, K. F., Bell, D. W., Wahrer, D. C. R., Schiripo, T. A., Haber, D. A., & Hariharan, I. K. (2002). salvador promotes both cell cycle exit and apoptosis in Drosophila and is mutated in human cancer cell lines. *Cell*, *110*(4), 467–478. https://doi.org/10.1016/S0092-8674(02)00824-3

Tashiro, E., Zako, T., Muto, H., Itoo, Y., Sörgjerd, K., Terada, N., ... Ariga, H. (2013). Prefoldin protects neuronal cells from polyglutamine toxicity by preventing aggregation formation. *Journal of Biological Chemistry*, *288*(27), 19958–19972. https://doi.org/10.1074/jbc.M113.477984

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V, ... Natale, D. A. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, *4*, 41. https://doi.org/10.1186/1471-2105-4-41

Tatusov, R. L., Koonin, E. V, & Lipman, D. J. (1997). A Genomic Perspective on Protein Families. *Science*, *278*(5338), 631–637. https://doi.org/10.1126/science.278.5338.631

Telerman, A., & Amson, R. (2009). Malignant Transformation, *9*(MARch). https://doi.org/10.1038/nrc2589

Terada, N., Takase, K., Papst, P., Nairn, a C., & Gelfand, E. W. (1995). Rapamycin inhibits ribosomal protein synthesis and induces G1 prolongation in mitogen-activated T lymphocytes. *Journal of Immunology (Baltimore, Md. : 1950)*, *155*(7), 3418–3426.

Terzi, M. Y., Izmirli, M., & Gogebakan, B. (2016). The cell fate: senescence or quiescence. *Molecular Biology Reports*. https://doi.org/10.1007/s11033-016-4065-0

Thaw, P., Baxter, N. J., Hounslow, A. M., Price, C., Waltho, J. P., & Craven, C. J. (2001). Structure of TCTP reveals unexpected relationship with guanine nucleotide-free chaperones. *Nature Structural Biology*, *8*(8), 701–704. https://doi.org/10.1038/90415

Thayanithy, V., & Venugopal, T. (2005). Evolution and expression of translationally

controlled tumour protein (TCTP) of fish. *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology*, *142*(1), 8–17. https://doi.org/10.1016/j.cbpc.2005.04.011

The Gene Ontology Consortium. (2017). Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Research*, *45*(D1), D331–D338. https://doi.org/10.1093/nar/gkw1108

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, *408*(6814), 796–815. https://doi.org/10.1038/35048692

Thiele, H., Berger, M., Lenzner, C., Kuhn, H., & Thiele, B. (1998). Structure of the promoter and complete sequence of the gene coding for the rabbit translationally controlled tumor protein (TCTP) P23. *European Journal of Biochemistry*, *257*(1), 62–68. https://doi.org/10.1046/j.1432-1327.1998.2570062.x

Thiele, H., Berger, M., Skalweit, A., & Thiele, B.-J. (2000). Expression of the gene and processed pseudogenes encoding the human and rabbit translationally controlled tumour protein (TCTP). *European Journal of Biochemistry*, *267*(17), 5473–5481. https://doi.org/10.1046/j.1432-1327.2000.01609.x

Thomas, G., Thomas, G., & Luther, H. (1981). Transcriptional and translational control of cytoplasmic proteins after serum stimulation of quiescent Swiss 3T3 cells. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(9), 5712–5716. https://doi.org/10.1073/pnas.78.9.5712

Thomas, N. S. B. (2004). Cell Cycle Regulation. In L. D. Griffin, J. D. Linch, D. C. Lowenberg, B. London, & M. Dunitz (Eds.), *Textbook of Malignant Haematology* (2nd ed., pp. 33–63). London.

Thomas, N. S., Pizzey, a R., Tiwari, S., Williams, C. D., & Yang, J. (1998). p130, p107, and pRb are differentially regulated in proliferating cells and during cell cycle arrest by alpha-interferon. *The Journal of Biological Chemistry*, *273*(37), 23659–23667. https://doi.org/10.1074/jbc.273.37.23659

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., ... Narechania, A. (2003). PANTHER: A library of protein families and subfamilies indexed by function. *Genome Research*, *13*, 2129–2141. https://doi.org/10.1101/gr.772403

Thompson, B. J., & Cohen, S. M. (2006). The Hippo Pathway Regulates the bantam microRNA to Control Cell Proliferation and Apoptosis in Drosophila. *Cell*, *126*(4), 767–774. https://doi.org/10.1016/j.cell.2006.07.013

Thomson, E., Ferreira-Cerca, S., & Hurt, E. (2013). Eukaryotic ribosome biogenesis at a glance. *Journal of Cell Science*, *126*(Pt 21), 4815–21. https://doi.org/10.1242/jcs.111948

Tibbetts, R. S., Brumbaugh, K. M., Williams, J. M., Sarkaria, J. N., Cliby, W. A., Shieh, S. Y., ... Abraham, R. T. (1999). A role for ATR in the DNA damage-induced phosphorylation of p53. *Genes & Development*, *13*(2), 152–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9925639

Trachana, K., Forslund, K., Larsson, T., Powell, S., Doerks, T., von Mering, C., & Bork, P. (2014). A phylogeny-based benchmarking test for orthology inference reveals the limitations of function-based validation. *PloS One*, *9*(11), e111122. https://doi.org/10.1371/journal.pone.0111122

Trachana, K., Larsson, T. A., Powell, S., Chen, W.-H., Doerks, T., Muller, J., & Bork,

P. (2011). Orthology prediction methods: a quality assessment using curated protein families. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *33*(10), 769–80. https://doi.org/10.1002/bies.201100062

Tsuchiya, H., Iseda, T., & Hino, O. (1996). Identification of a novel protein (VBP-1) binding to the von Hippel-Lindau (VHL) tumor suppressor gene product. *Cancer Research*, *56*(13), 2881–2885. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8674032

Tsuji, T., Ficarro, S. B., & Jiang, W. (2006). Essential role of phosphorylation of MCM2 by Cdc7/Dbf4 in the initiation of DNA replication in mammalian cells. *Molecular Biology of the Cell*, *17*(10), 4459–72. https://doi.org/10.1091/mbc.E06-03-0241

Tsukaya, H. (2013). Leaf development. *The Arabidopsis Book*, *11*, e0163. https://doi.org/10.1199/tab.0163

Tumaneng, K., Russell, R. C., & Guan, K. (2012). Organ Size Control by Hippo and TOR Pathways. *Current Biology*, *22*(9), R368–R379. https://doi.org/10.1016/j.cub.2012.03.003

Turner, J. J., Ewald, J. C., & Skotheim, J. M. (2012). Cell size control in yeast. *Current Biology*, *22*(9), R350-9. https://doi.org/10.1016/j.cub.2012.02.041

Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., ... Wilson, R. (2009). FlyBase: Enhancing Drosophila Gene Ontology annotations. *Nucleic Acids Research*, *37*(SUPPL. 1), 555–559. https://doi.org/10.1093/nar/gkn788

Tzfira, T., Li, J., Lacroix, B., & Citovsky, V. (2004). Agrobacterium T-DNA integration: Molecules and models. *Trends in Genetics*. https://doi.org/10.1016/j.tig.2004.06.004

Tzur, A., Kafri, R., LeBleu, V. S., Lahav, G., & Kirschner, M. W. (2009). Cell growth and size homeostasis in proliferating animal cells. *Science*, *325*(5937), 167–71. https://doi.org/10.1126/science.1174294

Tzur, A., Moore, J. K., Jorgensen, P., Shapiro, H. M., & Kirschner, M. W. (2011). Optimizing optical flow cytometry for cell volume-based sorting and analysis. *PloS One*, *6*(1), e16053. https://doi.org/10.1371/journal.pone.0016053

UniProt Consortium. (2014). Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Research*, *42*(Database issue), D191-8. https://doi.org/10.1093/nar/gkt1140

Usami, T., Horiguchi, G., Yano, S., & Tsukaya, H. (2009). The more and smaller cells mutants of Arabidopsis thaliana identify novel roles for SQUAMOSA PROMOTER BINDING PROTEIN-LIKE genes in the control of heteroblasty. *Development (Cambridge, England)*, *136*(6), 955–64. https://doi.org/10.1242/dev.028613

Vainberg, I. E., Lewis, S. A., Rommelaere, H., Ampe, C., Vandekerckhove, J., Klein, H. L., & Cowan, N. J. (1998). Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell*, *93*(5), 863–873. https://doi.org/10.1016/S0092-8674(00)81446-4

Van Auken, K., Fey, P., Berardini, T. Z., Dodson, R., Cooper, L., Li, D., ... Sternberg, P. W. (2012). Text mining in the biocuration workflow: applications for literature curation at WormBase, dictyBase and TAIR. *Database : The Journal of Biological*

Databases and Curation, *2012*, bas040. https://doi.org/10.1093/database/bas040

van den Heuvel, S., & Dyson, N. J. (2008). Conserved functions of the pRB and E2F families. *Nature Reviews. Molecular Cell Biology*, *9*(9), 713–724. https://doi.org/10.1038/nrm2469

van der Heijden, R. T., Snel, B., van Noort, V., & Huynen, M. A. (2007). Orthology prediction at scalable resolution by phylogenetic tree analysis. *BMC Bioinformatics*, *8*(1), 83. https://doi.org/10.1186/1471-2105-8-83

van der Windt, G. J. W., & Pearce, E. L. (2012). Metabolic swicthing and fuel choice during T-cell differentiation and memory development. *Immunological Reviews*, *249*(1), 27–42. https://doi.org/10.1111/j.1600-065X.2012.01150.x.Metabolic

Vickaryous, M. K., & Hall, B. K. (2006). Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biological Reviews of the Cambridge Philosophical Society*, *81*(3), 425–455. https://doi.org/10.1017/S1464793106007068

Vidal, A., & Koff, A. (2000). Cell-cycle inhibitors: Three families united by a common cause. *Gene*, *247*(1–2), 1–15. https://doi.org/10.1016/S0378-1119(00)00092-5

Walsh, J. B. (1995). How often do duplicated genes evolve new functions? *Genetics*, *139*(1), 421–428.

Wang, X., Fonseca, B. D., Tang, H., Liu, R., Elia, A., Clemens, M. J., ... Proud, C. G. (2008). Re-evaluating the roles of proposed modulators of mammalian target of rapamycin complex 1 (mTORC1) signaling. *Journal of Biological Chemistry*, *283*(45), 30482–30492. https://doi.org/10.1074/jbc.M803348200

Wang, Z. Q., Li, G. Z., Gong, Q. Q., Li, G. X., & Zheng, S. J. (2015). OsTCTP, encoding a translationally controlled tumor protein, plays an important role in mercury tolerance in rice. *BMC Plant Biology*, *15*, 123. https://doi.org/10.1186/s12870-015-0500-y

Weigmann, K., Cohen, S. M., & Lehner, C. F. (1997). Cell cycle progression, growth and patterning in imaginal discs despite inhibition of cell division after inactivation of Drosophila Cdc2 kinase. *Development (Cambridge, England)*, *124*(18), 3555–3563.

Weinberg, R. A. (1995). The retinoblastoma protein and cell cycle control. *Cell*, *81*(3), 323–330. https://doi.org/10.1016/0092-8674(95)90385-2

Weinkove, D., Neufeld, T. P., Twardzik, T., Waterfield, M. D., & Leevers, S. J. (1999). Regulation of imaginal disc cell size, cell number and organ size by Drosophila class I(A) phosphoinositide 3-kinase and its adaptor. *Current Biology*, *9*(18), 1019–1029. https://doi.org/10.1016/S0960-9822(99)80450-3

Weiss, A. (1993). T cell antigen receptor signal transduction: a tale of tails and cytoplasmic protein-tyrosine kinases. *Cell*, *73*(2), 209–12. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8477442

Weiss, R. L., Kukora, J. R., & Adams, J. (1975). The relationship between enzyme activity, cell geometry, and fitness in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, *72*(3), 794–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1093169

White, J. K., Gerdin, A. K., Karp, N. A., Ryder, E., Buljan, M., Bussell, J. N., ... Steel, K. P. (2013). XGenome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell*, *154*(2). https://doi.org/10.1016/j.cell.2013.06.022

Wieman, H. L., Wofford, J. A., & Rathmell, J. C. (2007). Cytokine stimulation promotes glucose uptake via phosphatidylinositol-3 kinase/Akt regulation of Glut1 activity and trafficking. *Molecular Biology of the Cell*, *18*(4), 1437–46. https://doi.org/10.1091/mbc.E06-07-0593

Williams, C. D., Linch, D. C., Sørensen, T. S., La Thangue, N. B., & Thomas, N. S. (1997). The predominant E2F complex in human primary haemopoietic cells and in AML blasts contains E2F-4, DP-1 and p130. *British Journal of Haematology*, *96*(4), 688–96. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9074408

Williams, G. H., Romanowski, P., Morris, L., Madine, M., Mills, A. D., Stoeber, K., ... Coleman, N. (1998). Improved cervical smear assessment using antibodies against proteins that regulate DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(25), 14932–7. https://doi.org/10.1073/pnas.95.25.14932

Williams, R. S., Shohet, R. V, & Stillman, B. (1997). A human protein related to yeast Cdc6p. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(1), 142–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8990175

Wintersberger, E., Rotheneder, H., Grabner, M., Beck, G., & Seiser, C. (1992). Regulation of thymidine kinase during growth, cell cycle and differentiation. *Advances in Enzyme Regulation*, *32*(C), 241–54. https://doi.org/10.1016/0065-2571(92)90020-Z

Woo, H. H., & Hawes, M. C. (1997). Cloning of genes whose expression is correlated with mitosis and localized in dividing cells in root caps of Pisum sativum L. *Plant Molecular Biology*, *35*(6), 1045–51. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9426627

Wood, V., Harris, M. a, McDowall, M. D., Rutherford, K., Vaughan, B. W., Staines, D. M., ... Oliver, S. G. (2012). PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Research*, *40*(Database issue), D695-9. https://doi.org/10.1093/nar/gkr853

Wullschleger, S., Loewith, R., & Hall, M. N. (2006). TOR signaling in growth and metabolism. *Cell*, *124*(3), 471–84. https://doi.org/10.1016/j.cell.2006.01.016

Wurmser, A. E., Sato, T. K., & Emr, S. D. (2000). New component of the vacuolar class C-Vps complex couples nucleotide exchange on the Ypt7 GTPase to SNARE-dependent docking and fusion. *The Journal of Cell Biology*, *151*(3), 551–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11062257

Xiao, B., Chen, D., Luo, S., Hao, W., Jing, F., Liu, T., ... Ma, Q. (2016). Extracellular translationally controlled tumor protein promotes colorectal cancer invasion and metastasis through Cdc42/JNK/ MMP9 signaling. *Oncotarget*, *7*(31), 50057–50073. https://doi.org/10.18632/oncotarget.10315

Xing, Y., & Lee, C. (2005). Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(38), 13526–31. https://doi.org/10.1073/pnas.0501213102

Xu, S., Zhang, Z., Jing, B., Gannon, P., Ding, J., Xu, F., ... Zhang, Y. (2011). Transportin-sr is required for proper splicing of resistance genes and plant immunity. *PLoS Genetics*, *7*(6), 1–10. https://doi.org/10.1371/journal.pgen.1002159

Yamamoto, K., Gandin, V., Sasaki, M., McCracken, S., Li, W., Silvester, J. L., ... Mak, T. W. (2014). Largen: a molecular regulator of mammalian cell size control. *Molecular Cell*, *53*(6), 904–15. https://doi.org/10.1016/j.molcel.2014.02.028

Yamamoto, K., & Mak, T. W. (2017). Mechanistic aspects of mammalian cell size control. *Development, Growth & Differentiation*, *59*(1), 33–40. https://doi.org/10.1111/dgd.12334

Yan, Z., DeGregori, J., Shohet, R., Leone, G., Stillman, B., Nevins, J. R., & Williams, R. S. (1998). Cdc6 is regulated by E2F and is essential for DNA replication in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(7), 3603–8. https://doi.org/10.1073/pnas.95.7.3603

Yang, K., & Chi, H. (2012). MTOR and metabolic pathways in T cell quiescence and functional activation. *Seminars in Immunology*, *24*(6), 421–428. https://doi.org/10.1016/j.smim.2012.12.004

Yang, K., & Chi, H. (2018). *Cellular Quiescence*. (H. D. Lacorazza, Ed.) (Vol. 1686). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-7371-2

Yang, K., Neale, G., Green, D. R., He, W., & Chi, H. (2011). The tumor suppressor Tsc1 enforces quiescence of naive T cells to promote immune homeostasis and function. *Nature Immunology*, *12*(9), 888–897. https://doi.org/10.1038/ni.2068

Yankulov, K., Todorov, I., Romanowski, P., Licatalosi, D., Cilli, K., McCracken, S., ... Bentley, D. L. (1999). MCM proteins are associated with RNA polymerase II holoenzyme. *Molecular and Cellular Biology*, *19*(9), 6154–63. https://doi.org/10.1128/MCB.19.9.6154

Yébenes, H., Mesa, P., Muñoz, I. G., Montoya, G., & Valpuesta, J. M. (2011). Chaperonins: two rings for folding. *Trends in Biochemical Sciences*, *36*(8), 424–32. https://doi.org/10.1016/j.tibs.2011.05.003

Yenofsky, R., Bergmann, I., & Brawerman, G. (1982). Messenger RNA species partially in a repressed state in mouse sarcoma ascites cells. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(19), 5876–80. https://doi.org/10.1073/pnas.79.19.5876

Yu, F. X., Zhao, B., & Guan, K. L. (2015). Hippo Pathway in Organ Size Control, Tissue Homeostasis, and Cancer. *Cell*, *163*(4), 811–828. https://doi.org/10.1016/j.cell.2015.10.044

Yubero, N., Esteso, G., Cardona, H., Morera, L., Garrido, J. J., & Barbancho, M. (2009). Molecular cloning, expression analysis and chromosome localization of the Tpt1 gene coding for the pig translationally controlled tumor protein (TCTP). *Molecular Biology Reports*, *36*(7), 1957–1965. https://doi.org/10.1007/s11033-008-9405-2

Yurinskaya, V., Aksenov, N., Moshkov, A., Model, M., Goryachaya, T., & Vereninov, A. (2017). A comparative study of U937 cell size changes during apoptosis initiation by flow cytometry, light scattering, water assay and electronic sizing. *Apoptosis*, *22*(10), 1287–1295. https://doi.org/10.1007/s10495-017-1406-y

Zetterberg, A., & Killander, D. (1965). Quantitative cytochemical studies on interphase growth. *Experimental Cell Research*, *38*(38), 272–284.

Zetterberg, A., & Larsson, O. (1985). Kinetic analysis of regulatory events in G1 leading to proliferation or quiescence of Swiss 3T3 cells. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(16), 5365–9. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/3860868

Zhai, Y., Li, N., Jiang, H., Huang, X., Gao, N., & Tye, B. K. (2017). Unique Roles of the Non-identical MCM Subunits in DNA Replication Licensing. *Molecular Cell*, *67*(2), 168–179. https://doi.org/10.1016/j.molcel.2017.06.016

Zhai, Y., & Tye, B.-K. (2017). Structure of the MCM2-7 Double Hexamer and Its Implications for the Mechanistic Functions of the Mcm2-7 Complex. *Advances in Experimental Medicine and Biology*, *1042*, 189–205. https://doi.org/10.1007/978-981-10-6955-0_9

Zhang, C. J., Zhou, J. X., Liu, J., Ma, Z. Y., Zhang, S. W., Dou, K., ... He, X. J. (2013). The splicing machinery promotes RNA-directed DNA methylation and transcriptional silencing in Arabidopsis. *EMBO Journal*, *32*(8), 1128–1140. https://doi.org/10.1038/emboj.2013.49

Zhang, J., de Toledo, S. M., Pandey, B. N., Guo, G., Pain, D., Li, H., & Azzam, E. I. (2012). Role of the translationally controlled tumor protein in DNA damage sensing and repair. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(16), E926-33. https://doi.org/10.1073/pnas.1106300109

Zhang, J., Schneider, C., Ottmers, L., Rodriguez, R., Day, A., Markwardt, J., & Schneider, B. L. (2002). Genomic Scale Mutant Hunt Identifies Cell Size Homeostasis Genes in S. cerevisiae. *Current Biology*, *12*(23), 1992–2001. https://doi.org/10.1016/S0960-9822(02)01305-2

Zhang, Y.-J., Dai, Q., Sun, D.-F., Xiong, H., Tian, X.-Q., Gao, F.-H., ... Fang, J.-Y. (2009). mTOR Signaling Pathway Is a Target for the Treatment of Colorectal Cancer. *Annals of Surgical Oncology*, *16*(9), 2617–2628. https://doi.org/10.1245/s10434-009-0555-9

Zhao, B., Li, L., Tumaneng, K., Wang, C. Y., & Guan, K. L. (2010). A coordinated phosphorylation by Lats and CK1 regulates YAP stability through SCFβ-TRCP. *Genes and Development*, *24*(1), 72–85. https://doi.org/10.1101/gad.1843810

Zhao, B., Ye, X., Yu, J., Li, L., Li, W., Li, S., ... Lin, J. D. (2008). TEAD mediates YAP-dependent gene induction and growth control TEAD mediates YAP-dependent gene induction and growth control. *Genes & Development*, 1962–1971. https://doi.org/10.1101/gad.1664408

Zheng, D., Ye, S., Wang, X., Zhang, Y., Yan, D., Cai, X., ... Li, J. (2017). Pre-RC Protein MCM7 depletion promotes mitotic exit by Inhibiting CDK1 activity. *Scientific Reports*, *7*(1), 2854. https://doi.org/10.1038/s41598-017-03148-3

Zwieb, C., Glotz, C., & Brimacombe, R. (1981). Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. *Nucleic Acids Research*, *9*(15), 3621–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7024918

6.2 List of Supplementary tables

Table 6.2-1. List of terms listed in GO_SLIM database

Table is correct as of 31st January, 2018.

Term ID	Term Name
GO:0000003	reproduction
GO:0000902	*cell morphogenesis*
GO:0002376	immune system process
GO:0003013	*circulatory system process*
GO:0005975	carbohydrate metabolic process
GO:0006091	*generation of precursor metabolites and energy*
GO:0006259	DNA metabolic process
GO:0006397	*mRNA processing*
GO:0006399	tRNA metabolic process
GO:0006412	*translation*
GO:0006457	protein folding
GO:0006461	*protein complex assembly*
GO:0006464	cellular protein modification process
GO:0006520	*cellular amino acid metabolic process*
GO:0006605	protein targeting
GO:0006629	*lipid metabolic process*
GO:0006790	sulphur compound metabolic process
GO:0006810	*transport*
GO:0006913	nucleocytoplasmic transport
GO:0006914	*autophagy*
GO:0006950	response to stress
GO:0007005	*mitochondrion organization*
GO:0007009	plasma membrane organization
GO:0007010	*cytoskeleton organization*
GO:0007034	vacuolar transport
GO:0007049	*cell cycle*
GO:0007059	chromosome segregation
GO:0007067	*mitotic nuclear division*
GO:0007155	cell adhesion
GO:0007165	*signal transduction*
GO:0007267	cell-cell signaling
GO:0007568	*aging*
GO:0008150	biological process
GO:0008219	*cell death*

GO:0008283	cell proliferation
GO:0009056	*catabolic process*
GO:0009058	biosynthetic process
GO:0009790	*embryo development*
GO:0015979	photosynthesis
GO:0016192	*vesicle-mediated transport*
GO:0019748	secondary metabolic process
GO:0021700	*developmental maturation*
GO:0022607	cellular component assembly
GO:0022618	*ribonucleoprotein complex assembly*
GO:0030154	cell differentiation
GO:0030198	*extracellular matrix organization*
GO:0030705	cytoskeleton-dependent intracellular transport
GO:0032196	*transposition*
GO:0034330	cell junction organization
GO:0034641	*cellular nitrogen compound metabolic process*
GO:0034655	nucleobase-containing compound catabolic process
GO:0040007	*growth*
GO:0040011	locomotion
GO:0042254	*ribosome biogenesis*
GO:0042592	homeostatic process
GO:0043473	*pigmentation*
GO:0044281	small molecule metabolic process
GO:0044403	*symbiosis, encompassing mutualism through parasitism*
GO:0048646	anatomical structure formation involved in morphogenesis
GO:0048856	*anatomical structure development*
GO:0048870	cell motility
GO:0050877	*neurological system process*
GO:0051186	cofactor metabolic process
GO:0051276	*chromosome organization*
GO:0051301	cell division
GO:0051604	*protein maturation*
GO:0055085	transmembrane transport
GO:0061024	*membrane organization*
GO:0065003	macromolecular complex assembly
GO:0071554	*cell wall organization or biogenesis*
GO:0071941	nitrogen cycle metabolic process